The PHILOSOPHICAL MAGAZINE

FIRST PUBLISHED IN 1798

. 43 SEVENTH SERIES No. 339 April, 1952

A Journal of Theoretical Experimental and Applied Physics

PRICE 15s. 0d.
Annual Subscription £8 0s. 0d. payable in advance

AND PUBLISHED BY TAYLOR & FRANCIS LTD., RED LION COURT, FLEET ST., LONDON, E.C.4.

ADVANCES IN PHYSICS

A QUARTERLY SUPPLEMENT OF THE PHILOSOPHICAL MAGAZINE

On 1st January, 1952, the first number of this new Quarterly Supplement to the Philosophical Magazine was published. The aim of this Supplement will be to give those interested in physics comprehensive and authoritative accounts of recent important developments. It is felt by the Editor that in view of the rapid advances in many branches of physics, scientists will welcome a journal devoted to articles of this type.

VOLUME 1 APRIL 1952 NUMBER 2

Subject to alteration up to the time of going to Press

PRICE *per part* 15/- *plus postage* **PRICE** *per annum* £2 15s. 0d. *post free*

Printed and Published by
TAYLOR & FRANCIS, LTD., RED LION COURT, FLEET ST., LONDON, E.C.4

XXXV. *The Growth of Cadmium Iodide Crystals: II—A Study of the Heights of Growth Steps on Cadmium Iodide*

By A. J. Forty
H. H. Wills Physical Laboratory, University of Bristol*

[Received January 9, 1952]

Summary

Several methods for measuring step-heights on crystal surfaces are described. The measurements of small growth steps (<500 A high) on cadmium iodide, by an internal interference method, show that the theory of formation of large screw dislocation groups, based on a 'buckle, followed by slip' hypothesis, is probably correct. Dislocation groups having Burgers' vectors equal to any integral number of minimal sandwich layer thicknesses can be expected, but even values occur with much greater frequency. It has been shown that the occurrence of 'odd' dislocation groups indicates polytypism in cadmium iodide crystals.

§1. Introduction

Observations of the growth of cadmium iodide crystals from aqueous solution (Forty 1951 a, 1952) have shown that the tabular crystals develop in thickness by the dislocation mechanism, proposed by Frank (1949). Unlike beryl (Griffin 1950, 1951) and paraffin wax (Dawson and Vand 1951), which grow from simple screw dislocations of unit Burgers' vector, the cadmium iodide crystals have growth centres based on *screw dislocation groups*, of large composite Burgers' vectors. This latter type of growth is very similar to that of carborundum (Verma 1951, Frank 1951), and the crystals might be expected to exhibit some form of polytypism. So far, there has been no direct evidence for this from x-ray structural analyses. Preliminary measurements of some step-heights have shown that steps of different growth hills are not identical. If, as Frank suggests (1951), the superstructures of polytypic substances are determined by the nature of the screw dislocation groups on which the crystals grow, the measurement of the height of a large number of growth steps should indicate whether cadmium iodide can be polytypic. In order to distinguish between growth layers of different thicknesses, a high degree of accuracy is required. This paper describes some of the methods which have been employed and the results obtained.

* Communicated by Professor N. F. Mott, F.R.S.

§2. Accuracy of Step Measurement Required

Cadmium iodide crystals have a layer structure consisting of extended close-packed (0001) planes of cadmium ions sandwiched between close-packed (0001) planes of iodine ions. The unit cell contains two such sandwiches, one rotated through 60 degrees relative to the other. The crystals grow by the spreading of composite layers across the (0001) faces. It is convenient to express the height of a growth step, or thickness of a growth layer, by the number, M, of minimal sandwich layers contained in it. By the mechanism of crystal growth, this number is integral. According to Bozorth (1922) the thickness of a minimal sandwich layer is $C_0=6{\cdot}84$ A. Therefore, to distinguish between the M and $(M+1)$ types of steps, their heights must be measured to within at least a half of the minimal sandwich layer thickness, or about 3 A.

§3. Methods of Step-height Measurement

In this investigation three methods have been employed: multiple-beam Fizeau interferometry, an internal interference method, and a method of metal shadowing.

The multiple-beam Fizeau technique (Tolansky 1948) is a useful method for precise measurements of step-heights in the range (100–500) A. However, it has been found to be generally unsuitable for the present application because the small crystals have a somewhat complex surface structure.

The second method has proved to be the most useful:

The Internal Interference Method

The crystals are nucleated between a microscope slide and cover slip, with the latter undermost so that the extremely thin, plane parallel plates grow on it. If these are examined in reflection with unfiltered mercury light, the microscopic images are beautifully coloured by the interference of light reflected from the two surfaces. Thin tablets having no surface structure are distinguished by a completely uniform tint; the extremely uniform tint before and after the appearance of surface structure indicates that the two surfaces are always parallel. Thin tablets with growth structure on both surfaces are recognized by two crossed systems of uniformly tinted bands (see fig. 5, Plate XI*: photographed in parallel, mercury green light). If a plate is initially seated on the cover slip, growth usually ceases on this surface and the only structure consists of radial steps, extending from the points of emergence of screw dislocation groups to the edges of the crystals. Growth on the emergence of screw dislocation groups on the opposite surface, in a region of higher supersaturation, produces spiral step-line structure. Fig. 6 (Plate XI) is an internal interferogram for such a crystal containing a single screw dislocation group. The magnitude of the discontinuity of tint across the radial step shows that the height of the radial step equals that of the spiral step. The interferogram for a growth hill

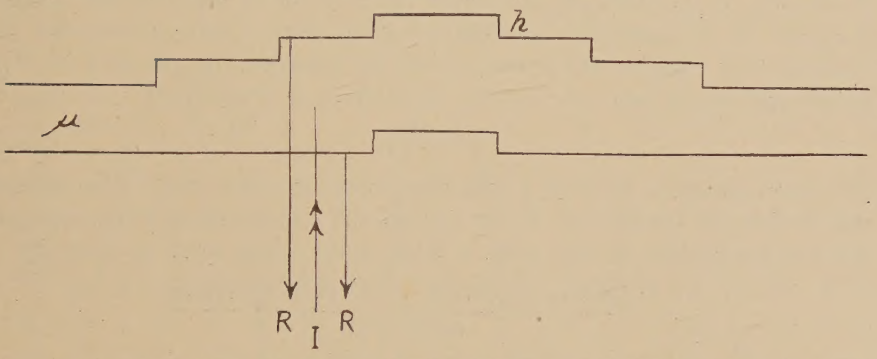
* For plates see end of issue.

centred on two screw dislocation groups of opposite hands consists of concentric bands of uniform tint (see fig. 7, Plate XI). There are two radial lines of discontinuity indicating the radial steps from the emergence of each screw dislocation group on the non-growing surface.

If the crystals with growth steps on one surface only are illuminated with parallel monochromatic light (mercury green, wavelength $\lambda=5461$ A has been employed principally), interference between reflections from the two surfaces produces a uniform shading corresponding to uniform crystal thickness. Fig. 1 is a schematic section through such a crystal plate. Neglecting absorption in the crystal and superposed interference effects due to the presence of other surfaces, the intensity of shade over an area where crystal thickness is t is represented by

$$I_t=A+B\cos\frac{4\pi\mu}{\lambda}.t,$$

Fig. 1



Schematic section through a crystal plate with a growth hill on one surface only.

where μ is the refractive index of the crystal in the [0001] direction. For a crystal with a growth hill of r equal steps on one surface only, the intensity of shade varies discontinuously over the hill as

$$I_t=A+B\cos\frac{4\pi\mu}{\lambda}.(t_0-rh), \quad \ldots\ldots\ldots \quad (1)$$

where h is the height of each step. The assumption that the growth hills provide a sequence of *equal steps* is confirmed later, in making the measurements (see Appendix).

Since r can only assume successive integral values, the intensity varies discontinuously across the step-lines in the image. If the growth hills were not stepped, that is if, in place of r, the thickness of the crystal depended on a continuous variable ρ, there would be maxima of intensity of shade at

$$\rho_{\max}=\frac{t_0}{h}-\frac{n\lambda}{2\mu h},$$

and minima at

$$\rho_{\min}=\frac{t_0}{h}-\frac{(2n+1)\lambda}{4\mu h},$$

n being an integer. The interval in ρ between successive maxima (or minima) would be

$$p=\frac{\lambda}{2\mu h}. \quad \ldots\ldots\ldots \quad (2)$$

The value of h can be evaluated accurately from this if p, which is not usually an integer, can be determined as accurately as possible from observations of intensity which are available only at integral values, r, of ρ. In the first approximation, p is the number of steps between one maximum and the next. If there are several maxima (or minima) p can be determined much more accurately than this by proper interpretation of the interferograms.

If a growth hill contains a large number of equal steps such that many maxima and minima of intensity are present, p can be determined accurately by visual examination. For this purpose a microphotograph is to be preferred to the direct microscopic image because the exaggeration of relation (1) by photography makes the location of maxima and minima sharper. It is generally possible to state that a true maximum (for continuously variable thickness) would lie between the rth and $(r+1)$th levels, and nearer, say, to the rth. If, then, it is located at

$$\rho=r+\tfrac{1}{4},$$

the worst possible error is $\frac{1}{4}$ and the probable error is $\frac{1}{8}$. The minima can be located similarly. If $3m+q$ ($q=0, \pm 1$) successive maxima and minima are located in this way, a fairly good value for p is given by

$$\frac{p}{2}=\frac{\Sigma(\text{last } m \text{ locations})-\Sigma(\text{first } m \text{ locations})}{m(2m+q)},$$

with an absolute probable error in p of

$$\frac{2\times\frac{1}{8}}{m^{1/2}(2m+q)}\sim\frac{1}{8m^{3/2}},$$

if m is large. This is better than taking the top and bottom halves $(3m/2, (3m/2)+q)$ of the locations, much better than simply taking the average value of p and not much worse than taking the moment, which is the best procedure. The probable error in the measurement of height of the growth steps is therefore approximately

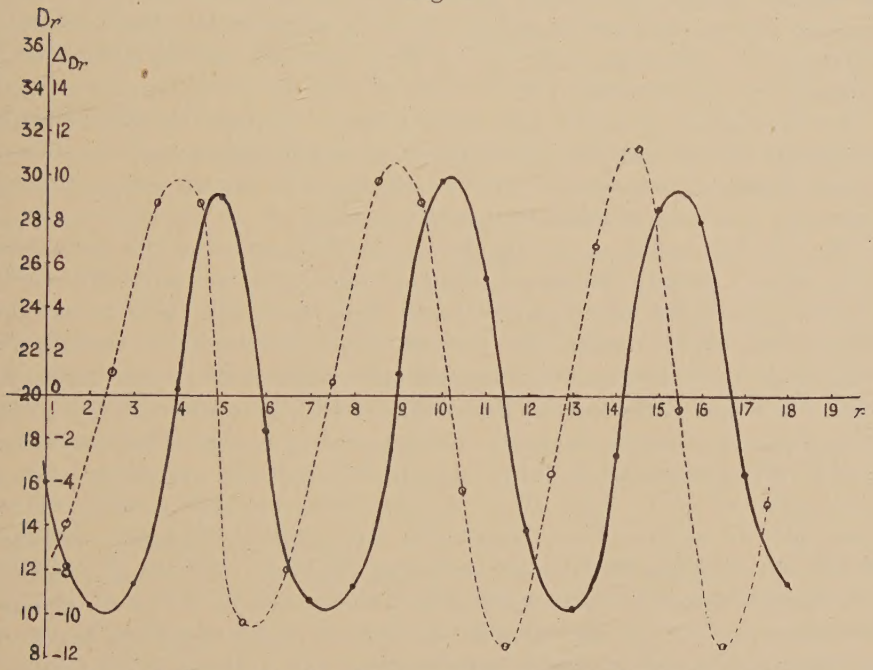
$$\frac{\lambda}{16\mu p^2 m^{3/2}}\sim\frac{200}{p^2 m^{3/2}}\text{ A.} \quad \ldots\ldots\ldots \quad (3)$$

For a growth hill with steps of 500 A, having $p\sim 3$, the probable error in the determination is about 3 A if $3m\sim 12$. In favourable circumstances some of the maxima and minima can be located with a probable error of less than 1/16, from the symmetrical distribution of intensities on either side. These can be suitably weighted when used in the evaluation of p.

For growth hills containing comparatively few steps and few maxima and minima of intensity, p is best determined by a graphical method,

From a microphotometer trace of photographic density across a micrograph of a growth hill, the density of each area between step-lines can be estimated. Denoting each area, from the centre outwards, 1, 2, 3, . . . r . . . , a graph of photographic density versus r is constructed (see fig. 2). Since only the maxima and minima of the graph are important, and the variation of intensity of shade with crystal thickness is zero at these points, it is correct to assume that the maxima and minima of photographic density correspond to minima and maxima of shade, independently of the photographic blackening law. Such a graph usually allows each maximum and minimum to be located with a probable error of less than 0·1.

Fig. 2



(*a*) Graph of photographic density versus r for internal interferogram (continuous curve).

(*b*) Graph of successive differences of photographic density (broken curve).

A rather better procedure is to draw a graph of successive differences of grain densities as a function of r, and take zeros of this graph as marking the maxima and minima of density (see fig. 2). If the variation of grain density is relation (1) deformed by the photographic blackening law, then Grain density,

$$D_r = a_0 + a_2(r-\delta)^2 + a_4(r-\delta)^4 + \ldots + a_{2i}(r-\delta)^{2i} + \ldots ,$$

when a true maximum or minimum occurs at $(n+\delta)$ and there are readings, D_r, at $(n+r)$, $r=0, \pm 1, \pm 2 \ldots$. The graph of differences is given by

$$\Delta_{r+1/2} = D_{r+1} - D_r = \Sigma a_{2i}[(r-\delta+1)^{2i} - (r-\delta)^{2i}],$$

and this, regarded as a continuous function of r, is zero when

$$(r-\delta)=-\tfrac{1}{2}, \quad (r-\delta+1)=+\tfrac{1}{2},$$

i.e. when $$r+\tfrac{1}{2}=\delta.$$

Again, the maxima and minima of light intensity can be located from a micrograph, independently of the photographic blackening law. From the graph of differences, the maxima and minima can now be located with a probable error of less than 0·05. Proceeding as before with $3m+q$ locations of maxima and minima, p can be determined with an absolute probable error of $0{\cdot}05/m^{3/2}$. The corresponding probable error in the step-height is about $80/m^{3/2}\,.\,p^2$ A so that a 500 A ($p\sim3$) step determined in this way has a probable error of about 1 A if about 12 maxima and minima are available. A 100 A step can be measured from a growth hill containing only two or three maxima and minima with a probable error of less than 1 A.

The evaluation of step-height from relation (2) requires an accurate knowledge of λ, μ and p. The value of p can be determined accurately by the methods outlined at length above, and λ is known with sufficient accuracy. μ is a more difficult quantity to find.

The crystals are strongly negatively birefringent. ω, the refractive index in the [0001] direction, which is the value required for these measurements (parallel, monochromatic light normally incident on the (0001) surfaces), is outside the present range of immersion techniques. Preliminary estimations by the immersion method show that ω is greater than 1·7, whilst a theoretical estimation, from the atomic refractivities of cadmium and iodine, suggests that ω is less than 2·0. One particular growth hill was found to have $p=27\pm0{\cdot}2$; which corresponds to a step-height of $59{\cdot}7\pm0{\cdot}4$ A if $\omega=1{\cdot}7$ or $50{\cdot}6\pm0{\cdot}4$ A if $\omega=2{\cdot}0$. It is reasonable to assume that the growth steps will be integral multiples of the minimal sandwich layer thickness (from the nature of growth), and the only integral step between these two limits is $8\times6{\cdot}84$ A or 54·72 A. Therefore $p=27\pm0{\cdot}2$ corresponds to a step of eight minimum layers of cadmium iodide. From these considerations, the value of ω for mercury green light ($\lambda=5461$ A) is estimated to be $1{\cdot}850\pm0{\cdot}15$. Using this value of ω, the number of minimal sandwich layers, M, in steps of other growth hills can be calculated. If p for these growth hills is known very accurately, then a more accurate value of ω can be obtained by correcting M to a whole number. By averaging these more accurate values, the final value is found to be $\omega=1{\cdot}849\pm0{\cdot}001$. This determination of the refractive index of cadmium iodide has been checked by measuring the height of steps of a particular hill by the Fizeau multiple-beam interferometry, using reflections from the crystal surface and a glass reference flat, and by the internal interference method. The excellent agreement between the two methods confirms this value of ω.

This method of internal interference has many advantages over the other techniques. It is a simple precision method for measuring steps

on a *growing* crystal, with a wide range of applicability. The lower limit is fixed at about 50 A ($p \sim 28$) below which the steps cannot be counted. There is an upper limit in the neighbourhood of $\lambda/2\mu$ (1500 A) where the method becomes less simple because of ambiguities; the ambiguities can be removed by using two or more wavelengths. Apart from this there is an indefinite upper limit, dependent on the degree of accuracy required. Neglecting errors in λ and μ, the probable error in a step-height is normally $\langle 200/(p^2 m^{3/2})$ A which is about $30h^{1/2}/N^{3/2}$ A, where N is the total number of steps to be counted, $3mp$.

Since this is essentially a two-beam interference phenomenon, the conditions are not as stringent as in multiple-beam interferometry. Strict parallelism is not so important so that a high power microscope objective can be used and all growth hills can be resolved; this is essential for step-counting. The two interfering waves need not have exactly the same amplitudes, but the reflectivities of the two surfaces are expected to be similar.

Apart from measurement, this form of illumination is useful for examining the structure of growth hills. The contrast for small steps is high, particularly in the areas between maxima and minima, where the variation of intensity with crystal thickness is large. Figs. 8, 9, 10 (Plate XII) are microphotographs, taken under similar conditions of exposure, of a dry, unsilvered crystal surface with bright field vertical illumination, reflection phase contrast, and the internal interference illumination. The smallest step measured on cadmium iodide is about 55 A, and the contrast is as good as that of a similar step under phase contrast illumination. Much smaller steps can be seen clearly on lead iodide crystals; the method might be employed for observation of monomolecular steps on other crystals.

Method of Metal-Shadowing

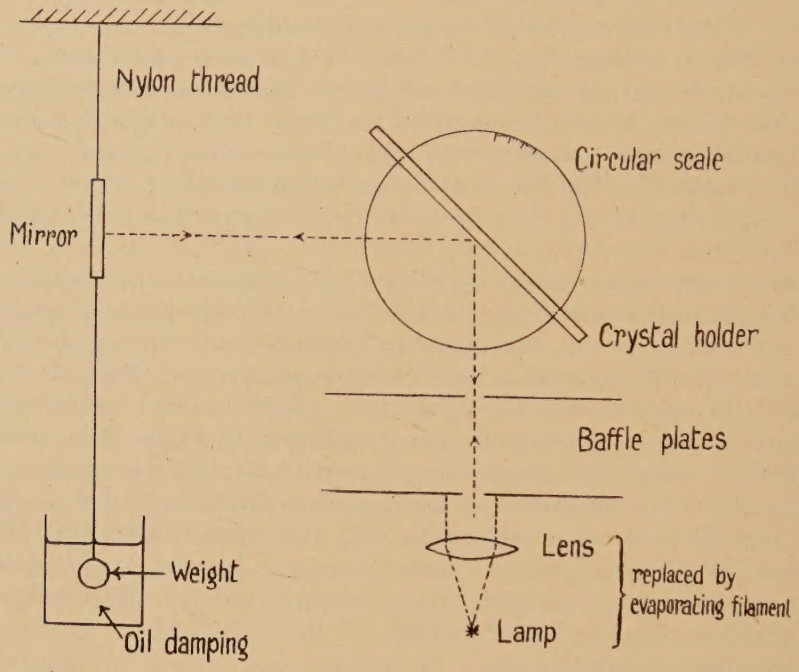
Some of the steps are very high (>1000 A). Therefore large shadows can be produced on the surfaces of dry crystals with a metal shadowing beam of low angle incidence, and measured under the optical microscope (see fig. 11, Plate XII). The comparatively low magnification of the latter necessitates an angle of about 5 degrees of arc between the (0001) surface and the shadowing beam. To attain the accuracy of ± 3 A in a 1000 A step, the angle of shadowing must be known to within 1 minute of arc. This high degree of accuracy demands a much more refined shadowing technique than that normally employed in the measurement of smaller steps by electron microscopy.

The shadowing beam is collimated to a narrow pencil of less than a minute of divergence and the angle of shadowing is set accurately to within a minute of arc by a circular scale oriented by an optical lever method, illustrated in principle in fig. 3.

The easiest metal to evaporate is silver. Palladium-gold gives sharper shadows but it is difficult to evaporate sufficient metal on the surfaces at 5 degrees of arc.

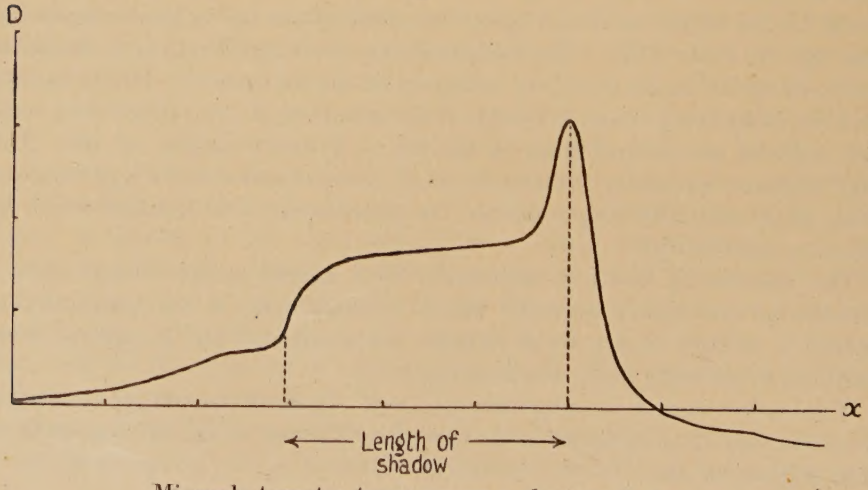
The lengths of the shadows are measured directly under the microscope by a micrometer screw in the eye-piece or from a microphotometer trace across a micrograph. The latter method is more accurate because allowance can be made for the non-sharpness of the edge of the shadow and for the superposed image of the step-line. Fig. 4 is a typical microphotometer trace across a photographed shadow.

Fig. 3



Schematic representation of orientation of crystal surface for metal shadowing.

Fig. 4



Microphotometer trace across a shadowed step.

The accuracy is not as high as anticipated, chiefly because the lengths of the shadows cannot be measured with sufficient accuracy. However, this is the best method for evaluating large steps in the range (1000–2000) A and the measurements made, accurate to within 20 A, are useful for indicating qualitatively the types of steps possible in this range.

The chief advantage of this technique over interferometric methods is that a step-height is unambiguously defined by the angle of shadowing and the length of the shadow in the direction of the shadowing beam.

§4. Discussion of Results Obtained

(*a*) The important results have been obtained for step-heights in the region (0–500) A by the internal interference method. About 200 determinations of step-heights of different growth hills have been made using this technique. Table 1 gives the distribution of growth steps on cadmium iodide in the range (0–350 A) (i.e. the distribution of numbers of unit layers, M, in the steps, in the range $0<M<50$). Measurements of step-heights greater than this are not always accurate enough for an exact determination of the integer M. Most of these determinations have been made by the visual method described above. The results show that all integral values of M are possible but that even values occur much more frequently than the odd values. Figs. 12 (*a*) and (*b*) are block diagrams used to examine the significance of the occurrence of odd values of M statistically. Fig. 12 (*a*) shows the frequency of occurrence of the differences of readings from the immediately previous even integer, blocked around 0·00, 0·25, 0·5, 0·75, 1·0, 1·25, 1·5, 1·75. Fig. 12 (*b*) shows the frequency of differences weighted according to the probable error in each reading, as calculated by formula (3), to increase the importance of the more accurate determinations. Some explanation of the preponderance of readings occurring exactly at integers, particularly the even integers, in table 1, although the average probable error is about 0·2, is required. M is determined from the relation,

$$M=\frac{\lambda}{2\mu\,.\,p\,.\,6{\cdot}84}=\frac{215{\cdot}9}{p},$$

for mercury green light. Since 215·9 is very close to $2^3\,.\,3^3$, M will be an exact integer whenever p is $2^n\,.\,3^m$ where $n,\ m\leqslant 3$. It is therefore not surprising that many of the interferograms have values of p which are indistinguishable from $2^n\,.\,3^m$ despite the fairly large calculated probable error, which is dependent on the number of steps counted for each determination. Figs. 12 (*a*), (*b*) are less conclusive than could be desired, and there remain two possible explanations of the values around the difference of 1·0. Such a distribution could be caused if most of the step-heights contain an even number of minimal sandwich layers with an average error of 0·2, and about 20% of the total contain an odd number of layers with the same average error. An equally possible, alternative explanation is that all of the step-heights contain an even number of layers but that about 20% of the measurements are in error

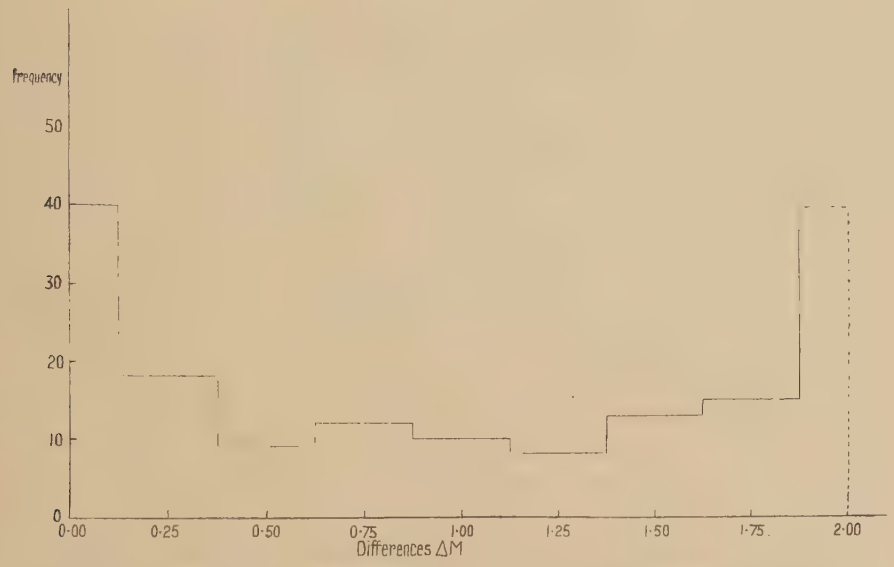
Table 1. Distribution of Step-heights on Cadmium Iodide Crystals, measured by Int
Interference Method. (Expressed in Multiples of 6·84 A)

1													
2													
3													
4													
5													
6													
7													
8	8·0 ±·08	8·0 ±·08											
9													
10													
11	11·35±·12	11·4 ±·1											
12	11·65±·2	12·0 ±·15	12·0 ±·2										
13													
14	13·95±·25	14·0 ±·15	14·25±·15	14·4 ±·5									
15													
16	16·0 ±·08	16·0 ±·1	16·0 ±·1	16·0 ±·1	16·0 ±·13	16·3±·2	16·35±·1						
17	16·8 ±·1*												
18	17·8 ±·06	17·75±·1	18·0 ±·05	18·0 ±·15	18·0 ±·15	18·0±·15	18·0 ±·15	18·0±·15	18·0 ±·2	18·0±·2	18·0±·2	18·0±·4	18
19	19·0 ±·15												
20	19·65±·25	19·65±·3	20·2 ±·2	20·3 ±·15	20·35±·25	20·45±·4	(20·5±·25)						
21	20·75±·25	20·75±·25	21·35±·3										
22	21·6 ±·2	21·6 ±·2	21·6 ±·2	21·6 ±·25	21·6 ±·4	21·7±·25	21·75±·15	21·8±·3	22·25±·25				
23	23·0 ±·2	23·0 ±·5											
24	23·8 ±·1	23·8 ±·2	24·0 ±·2	24·1 ±·12									
25	25·35±·2	25·4 ±·5											
26	(25·5±·2)	26·3 ±·25											
27	26·7 ±·35	27·0 ±·2*	27·0 ±·2*	27·0 ±·2	27·35±·25	27·4±·2							
28	28·0 ±·45	28·1 ±·07	28·1 ±·08	28·15±·1	28·2 ±·15	28·3±·1	28·3 ±·25	28·4±·3					
29	28·7 ±·5	28·75±·12	28·9 ±·2	29·1 ±·2	29·45±·4								
30	29·7 ±·15	29·7 ±·2	29·8 ±·5	29·9 ±·2	29·9 ±·2	30·0±·18	(30·5±·2)						
31	30·8 ±·2	31·3 ±·3	31·3 ±·6										
32	32·2 ±·15	32·2 ±·3											
33													
34	34·45±·25												
35	34·55±·2	34·7 ±·15											
36	35·9 ±·2*	35·95±·6	36·0 ±·01	36·0 ±·1	36·0 ±·12	36·0±·15	36·15±·15	36·4±·25					
37	37·2 ±·5												
38	(38·5±·5)												
39	39·0 ±·25												
40	39·6 ±·2	39·6 ±·45											
41	40·7 ±·2*	41·1 ±·2	41·4 ±·1	41·45±·15									
42	42·1 ±·12	42·1±·35*	42·2 ±·5										
43	42·8 ±·2*	42·8 ±·3	43·3 ±·5										
44													
45													
46													
47													
48	48·0±·2	48·0±·25											
49													
50													

* These step-heights have been re-measured by the more relia
microphotometric method of constructing a graph of succes
differences of photographic density, giving the values :—
17·0±·15, 26·65±·25, 26·7±·2, 35·9±·2, 40·9±·15, 42·1±·2, 43·35±

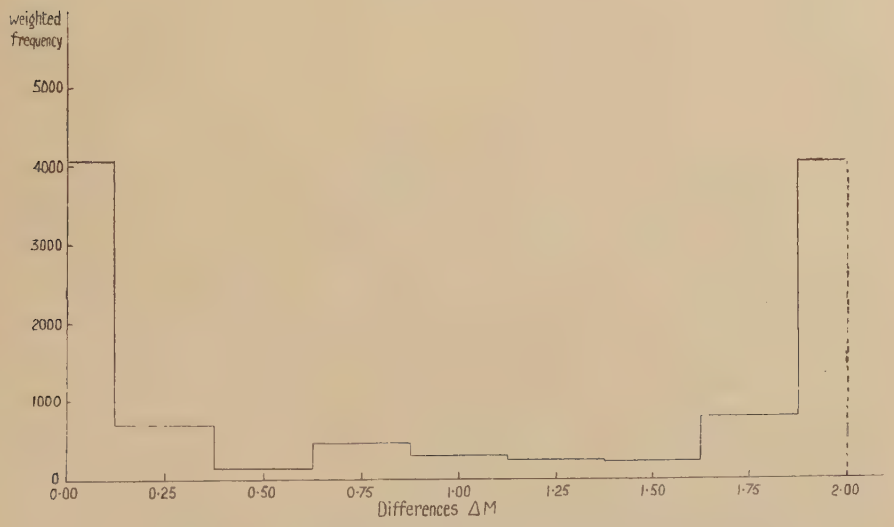
by more than is calculated statistically from the number of steps counted for each growth hill. The statistical analysis of the internal interferograms cannot take into account errors due to the superposition of interference patterns produced by other nearby surfaces or to the possible presence of additional unseen steps of small height. In order to decide which

Fig. 12 (*a*)



Block diagram of 'un-weighted' distribution of differences of M from previous even integer.

Fig. 12 (*b*)



Block diagram of 'weighted' distribution of differences of M from previous even integer.

of the above explanations of the distribution of step-heights is more probable, several of the step-heights corresponding to odd values of M with a small calculated probable error in table 1 have been checked by the more accurate graphical method from microphotometric readings. Some of the odd values were not confirmed. The value 42·8 gave 43·4 on re-measurement, and a reliable discrimination between 43 and 44 did not appear to be possible. On the other hand $16{\cdot}8 \pm 0{\cdot}1$ became $17{\cdot}0 \pm 0{\cdot}15$ and $40{\cdot}7 \pm 0{\cdot}2$ became $40{\cdot}9 \pm 0{\cdot}15$. It is therefore believed probable that some 10–20% of the step-heights genuinely correspond to an odd number of minimal sandwich layers. The Appendix shows the evaluation of a typical growth step containing an odd number of minimal layers. The step-height of the growth hill appears to be genuinely $M=41$ layers with a probable error of 0·15. The occurrence of odd values of M is important since it indicates the possibility of polytypism in cadmium iodide crystals.

The few measurements made by multiple-beam interferometry vary between 496 ± 10 A ($M=72$ minimum layers) and 1168 ± 10 A ($M=170$ minimum layers). The accuracy is not sufficient to distinguish between odd and even values of M.

The shadow measurements indicate that any step-height between 500 A and 3000 A is possible. Thus, qualitatively, M can have any integral value. In the preparation of shadowed specimens it has been observed that, in general, the steps of all the growth centres on one crystal face are of the same order of magnitude; that is they are all large (about 2000 A) or all small (about 500 A). This is to be expected if the ' buckle, followed by slip ' mechanism for production of dislocation groups (Frank 1951) is correct. For, if the crystal plate buckles fairly uniformly, the resultant slip will produce several screw dislocation groups of about the same size. The amount of buckling, and hence the size of the dislocation groups, will depend on the internal or external stresses.

(*b*) *Theory of Formation of Large Screw Dislocation Groups in Cadmium Iodide.*

The observation of the sudden appearance of spiral growth on the (0001) surfaces of the thin crystal plates has been described in Paper I. It is thought that the thin plates can be stressed beyond the theoretical yield stress, either internally by impurities or externally by convection currents etc. The resulting buckle is relieved by a slip of the cadmium iodide sandwiches past one another in the [0001] direction, with the formation of large screw dislocation groups, having their Burgers' vectors in this direction.

The stacking sequence of (0001) layers in the ideal cadmium iodide structure, as determined by Hägg (1948), can be described in the familiar a, b, c notation for sphere-packing lattices as

$$(a\gamma b)(c\alpha b)(a\gamma b)(c\alpha b)\ldots$$

where Roman letters represent (0001) close-packed layers of iodine ions, and Greek letters represent (0001) close-packed layers of cadmium ions.

Using the ' stacking operator ' notation of Nabarro and Frank (Frank 1951), the stacking is represented by :

$$\Delta'\Delta\nabla'\nabla\ldots$$

where the stacking operators define the relationship between successive layers of halogen ions and are primed where there is an intervening cation layer.

This is probably the structure of a thin plate in the initial stages of growth. After the buckling, parts of the plate, each having this ideal layer structure, slip past each other in the [0001] direction, to form screw dislocation groups. By such a mechanism there is an infinity of possible different dislocation groups (and, correspondingly, of growth steps on a surface when the plates begin to thicken) :

1. $(a\gamma b)(c\alpha b)(a\gamma b)(c\alpha b)\ldots$ $\longleftarrow$

slip in [0001] direction

$(a\gamma b)(c\alpha b)(a\gamma b)(c\alpha b)\ldots.$ $\longrightarrow$

If this slip, of one sandwich thickness, is terminated by a screw dislocation normal to (0001), the exposed step has the structure $\Delta'\nabla$. If the crystal grows on this, the repeating lattice unit will be $\Delta'\nabla$.

2. $(a\gamma b)(c\alpha b)(a\gamma b)(c\alpha b)\ldots$ $\longleftarrow$

slip

$(a\gamma b)(c\alpha b)(a\gamma b)(c\alpha b)\ldots.$ $\longrightarrow$

The exposed step structure is $\Delta'\Delta\nabla'\nabla$; this is also the repeating lattice unit if the crystal grows on the step.

3. $(a\gamma b)(c\alpha b)(a\gamma b)(c\alpha b)\ldots$ $\longleftarrow$

slip

$(a\gamma b)(c\alpha b)(a\gamma b)(c\alpha b)\ldots.$ $\longrightarrow$

The exposed step structure and, therefore, the repeating lattice unit after growth is $\Delta'\Delta\nabla'\nabla\Delta'\nabla$.

4. $(a\gamma b)(c\alpha b)(a\gamma b)(c\alpha b)\ldots$ $\longleftarrow$

slip

$(a\gamma b)(c\alpha b)(a\gamma b)(c\alpha b)\ldots$ $\longrightarrow$

The exposed step structure is $\Delta'\Delta\nabla'\nabla\Delta'\Delta\nabla'\nabla$, and the repeating lattice unit after growth is $\Delta'\Delta\nabla'\nabla$.

It follows that screw dislocation groups of two types should be formed in the crystal plates :—

(i) ' Even ' dislocation groups, having Burgers' vectors normal to the (0001) plane of $2n$ minimal sandwich layers, where $n=1, 2, 3\ldots$.

(ii) ' Odd ' dislocation groups having Burgers' vectors normal to the (0001) plane of $(2n+1)$ minimal sandwich layers, where $n=0, 1, 2, 3, \ldots$.

An even dislocation group is expected to be more stable, and hence occur more frequently, than an odd dislocation group. Assuming that the underlying structure is the ' fundamental ' one, found by Hägg, and described above, there is misfit on the surface of slip associated with the latter ; but there is perfect fit of the layers after slip of an even number of sandwiches. Since the step-height of a growth hill is determined by

the strength of the dislocation group (or exposed step terminating on the dislocation group), it follows that all integral values of M (number of minimum layer thicknesses contained in the growth steps) should be possible, but that the even values should occur more frequently than the odd values. This is exactly what has been found in the study of step-heights of growth hills.

(*c*) These results indicate that polytypism in cadmium iodide should be observed. According to this simple analysis, considering a single slip, x-ray investigations can only distinguish between structures built on the higher energy odd types of dislocation group, since all the even types will have essentially the same structure as the ideal crystal. Therefore, only the crystals growing on odd types of dislocation group will exhibit polytypism. A crystal growing on an odd dislocation group having Burgers' vector equal to five minimal sandwich layer thicknesses ($M=5$ for measurement of steps of the growth hill) should have a superstructure,

$$\Delta'\Delta\nabla'\nabla\Delta'\Delta\nabla'\nabla\Delta'\nabla.$$

There should thus be a superstructure in the [0001] direction of spacing equal to five minimal sandwich layer thicknesses. All the crystals having genuinely odd values of M in table 1 are polytypes of cadmium iodide.

Further evidence of the occurrence of polytypism in cadmium iodide crystals is given by fig. 27 of Paper I (Forty 1952). This is a micro-photograph of a growth hill illustrating the cross-lacing phenomenon. A similar phenomenon has been observed on type 6H silicon carbide crystals by Verma (1951) and has been explained satisfactorily by Frank (1951). Cross-lacing on a cadmium iodide crystal can only be understood if the stacking of the main growth layers contains two or more polytypic variants of the standard sequence resulting from growth on the emergence of a more complex dislocation group. Measurement of the thickness of the main growth layers of the pyramid by the internal interference method shows that M is $42{\cdot}8\pm0{\cdot}2$, which, according to the above classification, builds a polytype with a superstructure of this repeat in the [0001] direction.

It is conceivable that more complex 'odd' or 'even' dislocation groups can be formed if the slip mechanism takes place in a crystal plate already containing a polytypic variant in its layer structure. This is possible if further slip occurs in a crystal already thickening by growth from the emergence of an 'odd' dislocation group on the (0001) faces.

It is not surprising that such polytypism has not been found in x-ray analyses of cadmium iodide. A superstructure of axial spacing greater than 100 A is very difficult to detect, and the optical evidence shows that the crystals usually grow in layers much thicker than this. The filter paper technique for obtaining powder photographs, employed by Hägg (1948), will not reveal a superstructure, because the filter paper must be impregnated with a random selection of crystallites, grown from screw dislocation groups of different pitches. The resulting x-ray diffraction

pattern of a large number of different polytypes will not be distinguishable from that of the normal structure with a certain degree of random stacking disorder.

(*d*) Thus, the importance of the optical measurements of growth steps is two fold :

The distribution of step-heights of growth hills adds confirmation to the ' buckle, followed by slip ' hypothesis for the formation of large screw dislocation groups. The energy of misfit associated with slip of an odd number of sandwiches must be considerable, since odd values of M occur far less frequently than the even values. The step-height measurements indicate that large steps occur as frequently as the smaller steps. This is to be expected, since similar close-packed (0001) layers fit perfectly after slipping and the energy required to form a large, even dislocation group is not very different from that required to form a small group.

Secondly, the measurements give the first direct indication of polytypism in cadmium iodide crystals.

(*e*) An attempt is being made to repeat the optical measurements for the growth hills on lead iodide crystals, using the internal interference technique. Lead iodide has an ideal structure similar to that of cadmium iodide, but the stacking sequence found by Hägg (1948) is

$$(a\gamma b)(a\gamma b)\ldots$$

or

$$\Delta'\nabla.\ldots$$

Slip of any integral number of minimal sandwich layers should always result in a perfect fit in the (0001) planes. Therefore all dislocation groups (and hence growth steps) should be found with about the same frequency. Polytypism is not expected for this substance.

Acknowledgments

My thanks are due to Dr. F. C. Frank and Dr. J. W. Mitchell for their guidance and encouragement in this work and to the Department of Scientific and Industrial Research for a Maintenance Grant.

APPENDIX

Evaluation of step-height of growth hill from microphotometer trace illustrated in fig. 2.

Location of maxima and minima r	2·35	4·90	7·40	10·15	12·90	15·45
Successive differences Δr	2·55	2·50	2·75	2·75	2·55	

Probable error in each location of maxima and minima=0·05.

$$\Sigma(\text{last 2 locations})=28\cdot35,$$

$$\Sigma(\text{first 2 locations})=\ 7\cdot25,$$

$$p=\frac{28\cdot35-7\cdot25}{4}=5\cdot275\pm0\cdot017.$$

$$\text{Hence step-height}=\frac{2730}{1\cdot849\times5\cdot275}\text{ A}=279\cdot5(\pm1)\text{ A},$$

i. e. $M=40\cdot9\pm0\cdot15$ minimal sandwich layers.

Note.—The deviation of successive differences from the mean is 0·10, which is expected if the probable error in each location is 0·05. This, together with the regularity of the photographic density graphs (fig. 2), is evidence that the growth hill provides a sequence of many equal steps.

References

Bozorth, P. M., 1922, *J. Amer. Chem. Soc.*, **44,** 2232.
Dawson, I. M., and Vand, V., 1951, *Proc. Roy. Soc.* A, **206,** 555.
Forty, A. J., 1951 a, *Phil. Mag.*, **42,** 670 ; 1952, *Ibid.*, **43,** 72.
Frank, F. C., 1949, *Disc. Faraday Soc.*, No. 5 (Crystal Growth), 48 ; 1951, *Phil. Mag.*, **42,** 1014.
Griffin, L. J., 1950, *Phil. Mag.*, **41,** 196; 1951, *Ibid.*, **42,** 1337.
Hägg, G., 1948, *Colloques Internationaux du centre national de la recherche scientifique*, **10,** p. 5.
Tolansky, S., 1948, *Multiple-Beam Interferometry of Surfaces and Thin films*, (Oxford : University Press).
Verma, A. R., 1951, *Phil. Mag.*, **42,** 1005.

XXXVI. *The Waveforms of Atmospherics*

By P. G. F. Caton and E. T. Pierce
Cavendish Laboratory, Cambridge*

[Received December 20, 1951]

Summary

A convenient and economical technique for recording the waveforms of atmospherics is described, a time resolution of 3 microseconds being achieved with a photograph of total duration, 20 milliseconds. The character of the main waveform types observed at Cambridge has been examined with particular reference to the effects of the distance and direction of the originating lightning discharges. At night, the type of waveform observed depends upon the geographical location of the source; for storms to the south-west, a transition in waveform type occurs at a distance of about 1600 km; no similar change is observed for sources in a south-easterly direction. A critical discussion is given of the applicability of the theory of multiple ionospheric reflections to the various types of waveform, and of the accuracy with which the effective height of reflection and the distance of the source can be determined. Only a small proportion of atmospherics recorded, rarely more than 10% on any one night, yield precise values of these quantities; the height of reflection is found to be about 85 km. The smooth regular oscillatory type, originating beyond 1600 km to the south-west, cannot be explained by a straightforward application of the reflection theory.

§1. Introduction

The waveforms of 'atmospherics', the electromagnetic disturbances radiated by lightning discharges, show great diversity of character, and have been the subject of many previous studies. The important fact that certain waveforms could be interpreted as representing the results of multiple reflections between the ground and the ionosphere of a discrete pulse emitted by the source, was recognized by Laby, McNeill, Nicholls and Nickson (1940) in Australia. They derived values for the height of the reflecting layer from some atmospherics, but found that the majority of waveforms could not satisfactorily be explained on this simple mechanism. Schonland, Elder, Hodges, Phillips and van Wyk (1940) on the other hand, observing in S. Africa, considered that, at night, all atmospherics were explicable in terms of ionospheric reflections. In W. Europe, waveforms have been examined and classified by Lutkin (1939), and, more comprehensively, by Rivault (1945). Lutkin's classification refers predominantly to daytime and Rivault's to night-time conditions.

The present paper represents the results of an extensive series of observations conducted at Cambridge, England, from 1947 to 1951.

* Communicated by Dr. T. W. Wormell.

Both day and night conditions, all seasons of the year, and a wide range of distance of origin (100–4000 km) have been covered. Co-operation with the 'Sferics' organization of the British Meteorological Office (Ockenden 1947), which maintains a network of direction-finding stations for the accurate location of thunderstorm centres, has greatly enhanced the value of the work through the provision of numerous 'fixes' on individual atmospherics. A new and more complete classification of atmospheric waveforms has been derived and the behaviour of each variety studied in relation to the distance of origin; in addition, the surprising result has emerged that the type of waveform recorded depends upon direction as well as distance. A critical discussion is given of the applicability of the reflection theory to the different varieties of waveform. Difficulties and uncertainties, not previously considered, in the estimation of reflection heights and source distances from reflection type atmospherics are emphasized, and an assessment is made of the accuracy and reliability with which these parameters may be deduced from recorded waveforms.

§2. Experimental Details

The apparatus consists essentially of an aerial connected through a wide-band amplifier to a cathode-ray tube, together with a trigger device designed to select individual waveforms. A sphere (Wilson 1920) is employed as the aerial; its use upon an open site enables a reliable absolute measure of the amplitudes of atmospherics to be made. Any change in the vertical electric field E results in a corresponding voltage variation being applied to the amplifier. This is of conventional resistance-capacity coupled design with a band-width (to 3 db. points) from 50 c/s to 100 kc/s. The gain is variable from 300 to 30000, values suitable for radiation field strengths of 2 v/m and 20 mv/m respectively.

The presentation of the waveforms upon the tube is unusual. A linear time-base deflects horizontally, while another, with period an integral multiple of that of the horizontal sweep, is applied to one of the vertical deflector plates. Under undisturbed conditions a time-base 'raster' (usually five or nine horizontal sweeps) is visible upon the screen. The amplifier output is connected to the second plate deflecting vertically: any variations in field therefore produce corresponding vertical deflections superimposed upon the raster. Normally the brilliance of the tube is insufficient to record photographically, but upon the arrival of an atmospheric exceeding some predetermined amplitude, an electronic trigger automatically increases the brilliance for a time slightly less than the raster duration. The photographs start at some arbitrary point on the raster; each traverse is from left to right, and successive sweeps are displaced downwards. When the bottom right-hand corner is reached, the picture continues at the top left-hand corner. Positive fields deflect upwards.

The raster technique permits greater time resolution and duration of record than the orthodox single sweep, and is very economical in film

consumption. From a photograph of total duration 20 msec, time intervals can be measured with an estimated probable error of 3 μsec. With a tube voltage of 2·5 kv, an f/2 lens and Kodak R 55 film, successful photographs have been taken with writing speeds of 0·25 cm/μsec.

Auxiliary equipment, which may be employed if desired, includes a differentiating circuit giving an amplifier output proportional to dE/dt ; a delay line to restore the initial portion of the waveform, lost before triggering ; filters to remove long-wave stations ; and a compensator for stray 50 c/s fields radiated from nearby power lines. These additional circuits modify the overall response characteristics, but do not significantly distort the waveforms.

§ 3. The Classification of Atmospherics

In the present work, it has been found convenient to divide atmospherics by appearance into four main types. These are an irregular high-frequency variety, a regular peaked oscillation, a regular smooth oscillation, and those waveforms consisting of a long oscillatory train. The latter two types are separated into three and two sub-groups, respectively. Many atmospherics are transitional between two of the principal categories, while others, devoid of obvious characteristics, fall outside the classification.

The amplitudes, in the following description of the waveform types, are measured, unless otherwise specified, from zero to the largest peak. Since the use of the trigger involves a selection of the bigger waveforms, and because the Sferics network tend to fix only the larger disturbances, it is probable that the typical values of amplitude given exceed the true mean values, especially for the greater distances.

In the classification, no account has been taken of the ' slow tail ', the low frequency (100–700 c/s) component studied in detail by Watson-Watt, Herd and Lutkin (1937). A slow component accompanies some 15% of the regular peaked, regular smooth and long train waveforms, but is rarely associated with the irregular high-frequency type.

3.1. *The Irregular High-frequency Type.* (Plate XIII, figs. 2 and 3*)

At night, this type consists of an irregular series of high-frequency oscillations, peaks of the same sign being separated by intervals of 30 to 100 μsec. The fluctuations in amplitude are uneven, but there is generally a gradual decrease from the beginning to the end of the waveform. The duration may vary from 0·2 msec to 20 msec ; 1–3 msec is, however, the most frequent value. Typical amplitudes are 0·4 and 0·15 v/m for storms 250 and 700 km distant, respectively.

By day, the duration of these waveforms is in general rather less, the irregularities are perhaps not so pronounced, and the peaks appear more rounded.

These atmospherics correspond to Type I of Rivault (1945) and Group II of Lutkin (1939).

---

* For plates see end of issue.

3.2. *The Regular Peaked Type.* (Plate XIII, figs. 3 and 4 and Plate XIV, figs. 7 and 8)

This variety of atmospherics, observed only at night and during sunrise and sunset periods, consists of a characteristic oscillation with well-defined peaks. On waveforms of close origin, within some 400 km, successive pulses are usually separated by comparatively quiescent intervals, and the positive peaks are generally more pronounced (fig. 4). For distances greater than 400 km, the pulses forming the maxima and minima are normally of sufficient duration to produce an apparently continuous variation of field, at least in the early portion of the waveform (figs. 3 and 7). A characteristic change in the shape of the pulses with increasing order has frequently been noted, associated with doubling of the peaks of one sign, and accentuation of the negative peaks (fig. 12 (*b*)). Frequently one characteristic pulse shape (e.g. the ' sawtooth ' pulses on fig. 8), appears on the majority of waveforms observed from one storm centre during an observing period (10–15 min), and occasionally, two waveforms recorded more than a minute apart, and therefore from different discharge channels, have been observed identical in every detail.

The number of pulses discernible is greater at night than in the transitional periods at sunset and sunrise. The intervals between successive peaks of the same sign increase systematically from 50–150 μsec at the beginning of the waveform, to a final value, depending on the number of pulses visible, of 400–600 μsec. A typical form from a comparatively close discharge (*ca.* 300 km), soon after sunset, might show only four or five pulses, but waveforms observed later in the night from storms about 1 000 km distant frequently contain ten or more pulses, and have a total duration of 4–6 msec.

The amplitudes of individual waveforms from any one distance vary over a range of at least five or six to one. Values of 0·2–0·5 v/m appear to be typical both from storms 300 km distant (North Sea), and from storms at 1 000 km (N. Italy). At the closer distances, the first pulse is usually the largest, but beyond 600 km, the peak amplitude frequently corresponds to the second or third excursion. The amplitudes of the successive pulses on waveforms from sources about 1 000 km distant usually decrease rapidly from the second to about the eighth, and then comparatively slowly to the end of the waveform. A typical value for the tenth pulse is 25 mv/m.

The majority of waveforms of this type originate within 2 000 km of the observing station. They are frequently preceded by the irregular high-frequency type (fig. 3). The ratio of the amplitude of this to that of the main oscillation is very variable, but decreases with increasing distance of origin of the atmospheric. The initial excursion of the main oscillation is usually positive in sign, agreeing with the well-known predominance amongst flashes to ground of those carrying negative charge.

The corresponding types in the classifications of Rivault and Lutkin are Type 4 and Group I, respectively.

3.3. *The Regular Smooth Type*

3.3.1. *The 'Daytime' Type.* (Plate XIII, figs. 1 and 6 and Plate XIV, fig. 11)

The characteristic regular daytime waveform, from storms within 2000 km, consists of a few smooth oscillations rapidly diminishing in amplitude. The initial portion usually comprises a small positive peak followed by a large negative and then a large positive peak, the rapid decay in amplitude setting in thereafter. The number of oscillations increases with increasing distance of origin, averaging two or three for storms 500 km distant, and four to six for storms at 1500–2000 km. At greater distances, the oscillatory character is even more marked, and the third positive peak is frequently the largest. Typical amplitudes are 0·3 v/m and 0·1 v/m at 600 and 1800 km respectively. The quasi-period increases from about 100 μsec at the beginning, to about 220 μsec at the end of the waveform, and the total duration is usually 0·5–1·2 msec. A slight decrease in the mean quasi-period appears to accompany increased distance of origin, but due to the greater number of oscillations, waveforms from distant sources are of longer duration than those of close origin.

A high-frequency component frequently precedes this type of waveform, but at distances beyond 1500 km is generally of insufficient amplitude to initiate recording.

3.3.2. *The 'Quasi-sinusoidal' Type.* (Plate XIII, fig. 5 and Plate XIV, figs. 9 and 10)

This remarkable variety of atmospherics, observed only at night, consists of four to nine quasi-sinusoidal oscillations of large amplitude and gradually increasing period. Amplitudes of 0·15 v/m are typical from storms 2400 km distant. These waveforms are characterized, in contrast with the regular peaked type, by the smoothness of their oscillations (fig. 5), and the more gradual increase in the time intervals between the peaks. Slightly over 50% of the recorded examples, however, show a minor irregularity near the beginning of the waveform, the very smooth oscillations subsequently becoming re-established (fig. 9 (*a*)). The quasi-period increases steadily from 80–140 μsec at the beginning, to about 250 μsec at the end of the waveform. In most examples, the second or third oscillation is that of maximum amplitude, the quasi-period at this stage being 120–180 μsec. The amplitude of the oscillation decreases rapidly as the quasi-period passes 250 μsec.

All the recorded waveforms of this type, for which Sferics fixes are available, are from discharges beyond 1600 km. Few examples have been observed in which an irregular high-frequency component precedes a quasi-sinusoidal waveform, and these are generally for discharges within 2000 km. Any preceding irregular disturbance is therefore of small amplitude; probably less than 10% of the main oscillation for discharges at 2000 km, and proportionately smaller still at greater distances.

Quasi-sinusoidal waveforms have some similarities with oscillatory daytime forms from storms at comparable distances, but their amplitude is generally greater, the decrease in magnitude of the oscillations is more gradual, and the increase in period less rapid, factors contributing to the quasi-sinusoidal appearance. They have been observed at large amplitude from very great distances; on one occasion seven examples, from a localized storm-centre at approximately 4500 km, were of average amplitude 0·12 v/m.

This variety corresponds to Rivault's Type 3; Lutkin includes them within his Group I.

3.3.3. *Atmospherics of Very Small Amplitude.* (Plate XIII, figs. 1 and 5)

A second class of smooth waveform recorded at night contains three to five oscillations of amplitude usually less than 20 mv/m. The largest oscillation is normally the second, third or fourth, and the increase in quasi-period through the waveform is frequently more rapid than for the other smooth varieties. Typical values are 80 μsec at the beginning, and 250–300 μsec after five oscillations. These small waveforms, when recorded, occur by chance on a photograph initiated by one of the types previously described. They are believed to originate at 2000–4000 km, and, because of the more rapid increase in quasi-period along the waveform, seem to be essentially distinct from the quasi-sinusoidal variety.

A similar form, though with fewer oscillations, is observed by day. There is evidence that these atmospherics sometimes occur in bursts at intervals of a few milliseconds following a typical waveform of larger amplitude and distance of origin 500–2000 km.

3.4. *The Long Oscillatory Train Types*

Occasionally, at night, atmospherics are observed with up to forty or more oscillations, whose period increases steadily along the waveform to maximum values of 400 to 650 μsec according to the character of the peaks and the number of oscillations visible. Two kinds of pulse have been noted; a smooth type and a more peaked variety.

3.4.1. *The Peaked Long Train Type.* (Plate XIV, fig. 12 (*b*))

These waveforms seem to be an extension to twenty or more pulses of the regular peaked type previously described. A typical amplitude for the tenth order pulse is 40 mv/m.

3.4.2. *The Smooth Long Train Type.* (Plate XIV, fig. 12 (*a*))

These waveforms often appear as ripples superimposed upon a slow tail (amplitude *ca.* 0·1 v/m). The latter usually initiates the record (fig. 12 (*a*)), but on the rare occasions when the initial oscillations are registered, their peak amplitude is approximately 50 mv/m. The amplitude decreases rapidly after about the fourth peak, but a subsidiary maximum (*ca.* 10 mv/m) often occurs between the tenth and thirtieth orders. The increase in quasi-period along the waveform is slow but steady; the limiting period is often not attained by the end of the recorded waveform.

§4. Occurrence of the Different Waveform Types

4.1. *The Directional Effect*

During the period November 1950–February 1951, over 800 waveforms, fixed individually by the Sferics network were recorded; the observations covered both day and night conditions on over forty separate occasions. The atmospherics were classified, and the type of waveform observed related to the geographical location of the source.

These winter observations revealed the surprising fact that the variety of waveform recorded is dependent both upon the direction and the distance of the storm centre involved. A series of summer records taken in 1951 confirmed the findings of the previous winter in every respect, and it is reasonably certain, therefore, that the results of November 1950–February 1951, discussed in detail below, apply to all seasons of the year.

4.1.1. *Night-time Conditions*

Whenever a storm-centre was active at night either in Italy, Greece or the E. Mediterranean (i.e. to the S.E.), the atmospherics (20 nights, 220 fixes) were predominantly of the irregular high-frequency and regular peaked varieties ; not one quasi-sinusoidal waveform was observed from this direction. By contrast, 250 atmospherics, originating from storms 1 800–4 500 km distant over the Atlantic on 24 separate nights, were of the quasi-sinusoidal type. On several occasions, storms were active simultaneously over the Atlantic and in Italy, and, without exception, these produced quasi-sinusoidal and peaked waveforms respectively ; it is therefore improbable that any variations in propagation conditions from day to day could account for the observed distribution. The phenomenon is not solely an effect of increasing distance since atmospherics from storms 2 500 km to the south-east are predominantly of the regular peaked type, while those from only 2 000 km over the Atlantic are quasi-sinusoidal in character.

Storms were active over the Atlantic within 1 500 km on only four nights during the winter period under consideration ; most waveforms from these sources were of the high-frequency type, but some were of the regular peaked variety. In particular, on one evening (8th February 1951), storms were widespread in the Atlantic, and at 20.30 h GMT, there appeared to be a sharp transition between two regimes ; high-frequency and peaked types up to 1 400 km and the quasi-sinusoidal variety beyond 1 800 km. Observations at 21.30 h GMT showed that the transition zone had not moved. This eliminates the possibility that the observation of smooth waveforms from beyond 1 800 km was due to the ionosphere at this distance not having attained night-time conditions. Any such explanation is also disproved by the observation of quasi-sinusoidal atmospherics, from Atlantic storm-centres, throughout the night.

The locations given above were determined by the prevailing positions of the storm-centres. Other observations have shown that the irregular

high-frequency and regular peaked types are the predominant varieties observed from sources in the North Sea, France, Germany, the Netherlands, etc.

The observations have established that, for sources to the south-west, a transition from a peaked to a quasi-sinusoidal waveform occurs at a distance of about 1 600 km. No similar transition has been observed in a south-easterly direction. There is therefore a directional as well as a distance effect. This may originate at the source, arising either from meteorological influences (e.g. 'frontal' or 'heat' storms; storms in different air masses, Rivault 1950), or from a change in the discharge mechanism over land or sea; this latter explanation seems improbable since regular peaked atmospherics are observed both from European land sources and from the North and Mediterranean Seas. Alternatively, the variations in waveform may be due to differences in propagation over predominantly land or predominantly sea paths. It appears, however, that for the very long wavelengths involved, the electrical properties of land and sea are similar, and that the propagation characteristics are governed primarily by the comparatively poorly conducting ionosphere. Perhaps the most promising interpretation lies in considering the orientation of the path with respect to the earth's magnetic field.

Rivault (1948, 1950), observing from Poitiers, has also noted the predominance of smooth waveforms (his Type 3) from very distant maritime sources to the south-west, and of peaked atmospherics (his Type 4) from moderate distances to the south-east, but has claimed that the effect is entirely due to distance. He further subdivides his peaked waveforms into two varieties; one with narrow sharply defined peaks, and the other with broader more rounded peaks, the latter class being intermediate in appearance between his Types 3 and 4. This difference, although similar in kind to the first, he correlates with meteorological conditions at the source, and ascribes to differences in the lightning discharges from which the atmospherics originate.

4.1.2. *Day-time Conditions*

By day, only the irregular high-frequency and the regular smooth waveform types were observed. Records for the period November 1950–February 1951 (14 days, 180 fixes) indicated a definite tendency for the smooth waveforms to become more oscillatory with increasing distance. No pronounced directional effect was apparent, but smooth waveforms originating in the south-west generally contained more oscillations than those from storms at the same distance to the south-east (cp. Lutkin 1939). When two distinct storm-centres were active simultaneously, it was often possible to distinguish by appearance between the waveforms from the two centres.

4.2. *The Proportions of the Different Types Observed*

The use of a triggering system implies a certain selective effect, and it is impossible to give precise conclusions concerning the relative frequency

with which the various types of waveform occur, unless it is certain that the triggering mechanism responds equally to all the varieties of atmospherics encountered. Even if this is so, the typical waveform observed depends on the distance and direction of the storm-centres, and on whether day or night conditions are involved. Any distribution of atmospherics according to appearance, will therefore be controlled by the locations of the active storms, as well as by differences originating at the source; it will, in consequence, show wide variations.

Over 2000 atmospherics were recorded on sixteen evenings (18 h–21 h GMT) in October 1949. During this period, the main sources were situated to the south and east, but isolated storms were active over the Atlantic. Of the waveforms observed, 10% were of the high-frequency type, 30% of the regular peaked and associated long-train varieties, 20% of the smooth daytime and quasi-sinusoidal types, 4% atmospherics of small amplitude, while 36% could not satisfactorily be classed in any of the above categories. These figures are typical of many recording periods, but it must be emphasized that they are meaningless unless accompanied by specifications of the storm locations, season of the year, and time of observation.

§ 5. Interpretation of the Waveform Types

Photographic and field-change studies of the lightning flash have established the main features involved in a discharge to earth (Schonland, Malan and Collens 1935; Appleton and Chapman 1937; Schonland, Hodges and Collens 1938; Wormell and Pierce 1948). Beyond 300 km from the discharge, the field generated is primarily due to radiation. Since this is proportional to the rate of change of current in the flash, the major effects at great distances are produced by return strokes and any steps associated with leader processes.

The irregular high-frequency type probably originates in the steps of the initial leader to a discharge (Chapman 1939). A waveform of this variety followed by one of the regular peaked type, corresponds to a stepped leader succeeded by a return stroke; isolated high-frequency atmospherics could arise from 'air' or cloud discharges. Our observations show that high-frequency atmospherics followed, within the duration of the raster, by a regular waveform, are usually of duration 0·5–4 msec, whilst those of longer duration (5–20 msec) are either isolated or succeeded by only a comparatively small regular atmospheric.

The other three major waveform types are believed to be generated by return-strokes. (The bursts of small waveforms, which sometimes follow large daytime atmospherics (§ 3.3.3), may correspond to minor 'hook' components (Malan and Schonland 1947), succeeding the main return stroke.) The reflection mechanism (§ 1) is obviously effective, in certain instances, in modifying the original radiated pulse, and we consider that a critical discussion of the applicability of the reflection process in the interpretation of these three varieties of atmospheric, is

now desirable, especially in view of the unresolved discrepancies between the work of Laby *et al.* (1940) and Schonland *et al.* (1940) (§ 1).

5.1. *Theory of Multiple Ionospheric Reflections*

The mechanism of successive reflections between the ground and an ionospheric layer of constant height h, leads to the following formulae :—

$$ct_g = d, \quad \ldots\ldots\ldots \quad (1)$$

$$ct_n = \sqrt{(D^2+4n^2h^2)}, \quad \ldots\ldots\ldots \quad (2)$$

where t_g = Time interval between the emission of the primary pulse by the discharge and the arrival at the receiver of the ground wave.

t_n = Time interval between the emission of the primary pulse and the arrival at the receiver of the wave that has undergone n reflections at the ionosphere. (The pulse of nth order.)

c = Velocity of propagation of the disturbances (assumed equal to that of light).

d = Great circle distance on the earth between source and receiver.

R = Radius of the earth.

$D = d\sqrt{(1+h/R)}$.

Formula (2) is valid providing $d^2/(48n^2R^2) \ll h/R$, a condition usually fulfilled in practice. Other expressions which may be derived from (2) include

$$h = \frac{c}{2}\sqrt{\frac{T_1T_2(T_1+T_2)}{(r^2-q^2)T_1-(q^2-p^2)T_2}}, \quad \ldots\ldots \quad (3)$$

$$D^2 = \left(\frac{4h^2(q^2-p^2)-c^2T_1^2}{2cT_1}\right)^2 - 4p^2h^2, \quad \ldots\ldots \quad (4)$$

where p, q and r are the orders of three sky pulses, and

$$T_1 = t_q - t_p, \quad T_2 = t_r - t_q.$$

The time interval, $t_n - t_{n-1}$, between successive ionospheric pulses, ranges through the values

$$2h^2(2n-1)/(cD) \text{ for } D \gg 2nh, \quad h\sqrt{2}/c \text{ for } D = 2nh,$$

to

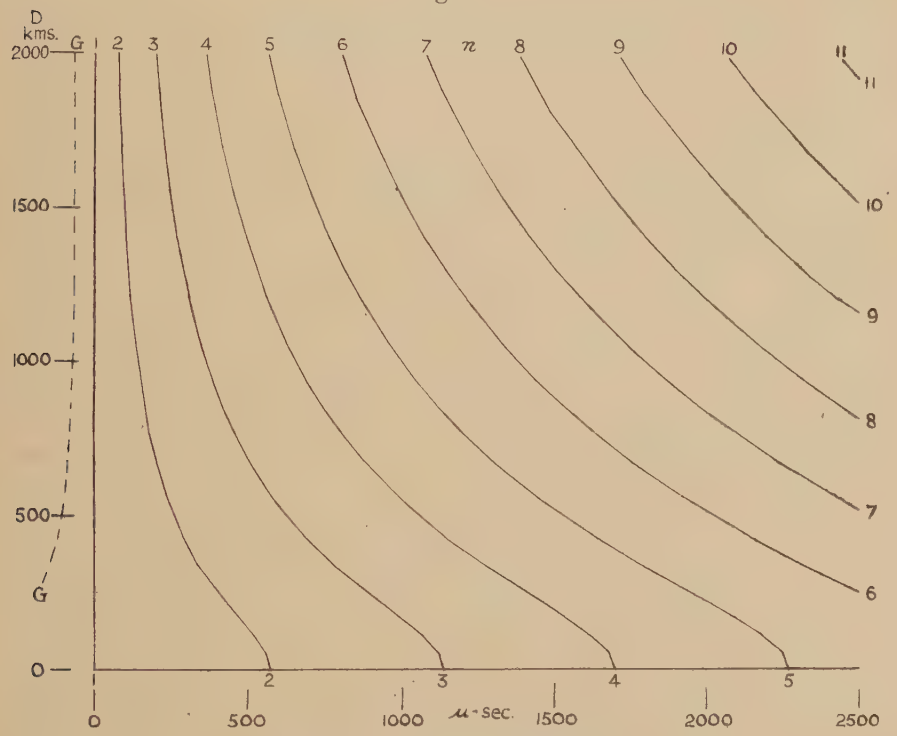
$$2h/c - D^2/\{4cn(n-1)h\} \text{ for } D \ll 2nh.$$

5.2. *Method of Analysis*

For evaluating D and h from the observed waveforms a graphical method of analysis has been found the most reliable and convenient. Two sets of graphs, based on expressions (1) and (2) have been constructed : (*a*) of time intervals, between the ground and successive sky pulses, against D, for constant values of h (fig. 13), and (*b*) of similar time intervals against h for constant values of D. A reference pulse is selected on the waveform and the time intervals between this and succeeding peaks of the same sign are measured. Trial values are then adopted for h and the reference peak order, these being varied until a constant distance corresponding to all the measured time intervals is derived from graphs (*a*). This, if obtainable, represents the solution

on the reflection mechanism, and is, effectively, a mean determination over the waveform. Reanalysis, using graphs (*b*), then indicates whether there is any variation, systematic or otherwise, in reflection height with order. It is often found that one or two peaks, not obviously identifiable by inspection, depart from the theoretical positions, corresponding to the average solution with constant h, by more than the errors of measurement; the present graphical method has the great advantage that such 'faulty' peaks are soon identified, and cannot therefore invalidate the complete analysis.

Fig. 13



h=85 km. Time intervals t_n–t_1.

Other methods of analysis have been given by Schonland *et al.* (1940) and Rivault (1943). Both involve the arbitrary selection of two or more sky pulses, and the application of formulae such as (3) and (4) (done graphically by Rivault), and suffer from the drawback that the selection of any 'faulty' peak giving completely false values for D and h, is not necessarily revealed by the method of analysis.

§6. Analysis of Individual Waveforms on the Reflection Mechanism

When the measured time intervals are compared with those predicted by the reflection theory, owing to the fact that there are two adjustable parameters, D and h, a range of possible solutions exists, and the results of the comparison are, therefore, often disappointingly inconclusive.

The difficulty is illustrated in table 1, where calculated time intervals between the ground wave and successive sky waves are given for three pairs of values of D and h.

Table 1. Calculated Time Intervals, $t_n - t_g$, in μsec

h (km)	D (km)	Order of sky-wave (n)							
		1	2	3	4	5	6	7	8
75	855	61	187	388	649	958	1306	1682	2080
80	1000	63	187	385	645	956	1308	1693	2102
85	1150	66	188	383	643	956	1311	1703	2122

When the peaks are sharply defined, the intervals can be measured with an accuracy of ± 3 μsec. Thus it would be impossible to distinguish between the three solutions tabulated unless seven or more reflected pulses were recorded. Usually, however, the peaks are broad, and doubt exists as to the exact points between which to measure. The errors of measurement may then approach ± 10 μsec, increasing still further the uncertainty of the solution. Even if an independent estimate of D is available from a Sferics fix, it is not always sufficiently accurate to improve appreciably the estimate of h. The most reliable analyses are obtained for waveforms which (*a*) originate within some 500 km, or (*b*) contain a large number of pulses.

6.1. *The Irregular High-frequency Type*

Due to the complicated nature of the original impulse, possible reflection effects in waveforms of this type are overshadowed, and they are unsuitable for analysis on the reflection theory.

6.2. *The Regular Peaked Type*

Several hundred waveforms of this type have been examined by the method of § 5.2. The separation of the pulses agree, in general, with the predictions of the reflection theory. In many cases, however, some peaks show small random displacements, about 20 μsec, from the positions predicted by the best average solution for the whole waveform. These deviations, which correspond to variations in D of 2–5%, exceed the errors of measurement. The first few pulses may also be displaced from their expected positions relative to the rest of the waveform; this may be caused by phase changes on reflection at the ionosphere.

The better fit with the theory is usually obtained by measuring the negative peaks for discharges beyond 400–500 km, and the positive pulses for waveforms of closer origin. This procedure implies, in general, the use of the more pronounced peaks. For storms beyond 500 km, the initial positive excursion appears to correspond to the ground wave, and the first negative peak to the first sky wave (cp. Chapman 1950): there is no obvious explanation for this empirical result.

Fits with the reflection theory have been classified as (A), (B), (C) or (D), according to the number of pulses available for measurement, and the consistency of individual determinations of D. It is estimated that (A) class analyses are accurate to ± 2 km in h, and $\pm 6\%$ in D. Corresponding figures for a (B) fit are ± 5 km in h, and $\pm 15\%$ in D, and for a (C) fit, ± 10 km and $\pm 30\%$, respectively. A (D) analysis implies that no significant value can be assigned to h or D. Of 200 peaked waveforms recorded in October 1949, 8%, 29% and 46% gave (A), (B) and (C) fits, respectively.

The distance deduced by analysis was compared wherever possible with the corresponding Sferics fix. In general, the better the fit, the better the agreement between the two determinations, but in practically all cases the difference was less than the combined errors. It is probable that the accuracy of some (A) class analyses exceeds that attained by direction-finding techniques.

The mean value derived from (A) analyses of the reflection height at night is 86 km. This agrees well with the 80–85 km, and the 85–90 km found by Laby *et al.* (1940) and Schonland *et al.* (1940), respectively. Rivault (1950) has given a value of 75 km. The present observations cannot be reconciled with the latter figure, since when sufficient reflections are visible, the time intervals between successive later order pulses usually exceed 500 μsec, the limiting value for $h=75$ km.

The great majority of regular peaked waveforms yield values of h between 85 and 90 km. No systematic variations for different conditions e.g. location of source, time of observation, season of year, etc., have been established, with the exception that waveforms recorded near sunset frequently indicate heights of 75 or 80 km.

Normally the reflection height is constant for all pulses, but occasionally some waveforms show a small increase with increasing order corresponding to more nearly vertical incidence on the ionosphere. Thus on one night 12 of the 17 atmospherics giving (A) or (B) analyses indicated a rise of 1–2 km in h for the higher orders, the mean value being 85 km.

It has been noted that the average quality of the analyses varies appreciably from one evening to another. In addition, on certain evenings, waveforms were observed, showing ionospheric reflections whose separation tended towards a limit of only 300 μsec (cp. Rivault Type 5). These gave approximate fits with $h=45$ km and $D=100$–400 km. Since there was no evidence for storms within 1 500 km, the fits must be considered fortuitous. No special conditions e.g. magnetic activity, have been correlated with the occurrence of these waveforms.

6.3. *The Regular Smooth Type*

Application of the reflection theory to 70 quasi-sinusoidal waveforms reveals that an approximate fit can sometimes be obtained if the first visible peak is assumed to correspond to a sky pulse of sufficiently high order (usually the 3rd, 4th or 5th). Even so, the deduced distances are not in agreement with Sferics fixes on individual examples, and it is

concluded that interpretation on a simple ray picture with peaks corresponding to sky pulses, is not valid for these waveforms. Budden (1951) has shown how propagation between an infinitely conducting earth and an ionospheric layer of finite conductivity, may transform a pulse into a smooth oscillation. This is perhaps the most promising approach to the explanation of smooth atmospherics.

Daytime waveforms, also do not give satisfactory results on the reflection mechanism.

Although the errors involved are large, the waveforms of small amplitude observed at night appear capable of interpretation by the reflection theory. Distances of 2000–4000 km and heights of 75–90 km are indicated, but no Sferics check is available due to the small amplitude.

6.4. *The Long Oscillatory Train Type*

These are explicable in terms of ionospheric reflections. The height of reflection ranges from 80 to 95 km, and the values of distance are 1000–4000 km for the peaked variety and 4000–9000 km for the smooth type.

§7. Conclusion

In this paper, atmospherics have been classified into four main types ; these categories represent broad divisions, and no extensive sub-grouping, such as is obviously possible, has been attempted. It has been shown that the nature of the atmospheric recorded, depends upon the distance and direction of the originating lightning discharge, but the precise explanation of the manner in which these factors combine to control the character of the waveform, remains obscure.

The detailed discussion of the simple ray theory of multiple ionospheric reflections indicates that this mechanism provides a satisfactory interpretation of the regular peaked and long oscillatory train varieties of waveform, but fails, in general, for the regular smooth type. It has been stressed that, even for the most favourable peaked waveforms, considerable practical difficulties are inherent in the determination of the height of the reflecting layer and the distance of origin. This latter quantity, however, can be estimated with an accuracy exceeding that obtained by direction-finding methods for a small and variable proportion (rarely exceeding 10%), of the total number of atmospherics, recorded on any one night.

It is hoped to publish further accounts of some specific types later.

Acknowledgments

The authors wish to express their gratitude to Dr. T. W. Wormell for his constant advice and encouragement both during the observations and in the preparation of this paper.

It is a pleasure to thank the Director of the Meteorological Office and the ' Sferics ' staff at the Central Forecasting Office, Dunstable, for their courtesy and co-operation in supplying Sferics results.

We are grateful to the Director of the Observatories, Cambridge University, for providing facilities for these investigations, and to Mr. F. Hepburn for assisting with the observations.

This work forms part of a programme of research supported by the Department of Scientific and Industrial Research. One of us (P. G. F. C.) wishes to thank the Department for his tenure of a maintenance grant.

References

Appleton, E. V., and Chapman, F. W., 1937, *Proc. Roy. Soc.* A, **158,** 1.
Budden, K. G., 1951, *Phil. Mag.*, **42,** 1.
Chapman, F. W., 1939, *Proc. Phys. Soc.*, **51,** 876 ; 1950. *U.R.S.I. Document A.G.* 1950/Nr 253/Comm IV.
Laby, T. H., McNeill, J. J., Nicholls, F. G., and Nickson, A. F. B., 1940, *Proc. Roy. Soc.* A, **174,** 145.
Lutkin, F. E., 1939, *Proc. Roy. Soc.* A, **171,** 285.
Malan, D. J., and Schonland, B. F. J., 1947, *Proc. Roy. Soc.* A, **191,** 485.
Ockenden, C. V., 1947, *Met. Mag.*, **76,** 78.
Rivault, R., 1943, *Note Prél. du L.N.R.* no. 32 ; 1945, *C.R. Acad. Sci., Paris,* **221,** 540 ; 1948, *Ibid.*, **226,** 1300 ; 1950, *Ibid.*, **230,** 1846.
Schonland, B. F. J., Elder, J. S., Hodges, D. B., Phillips, W. E., and van Wyk, J. W., 1940, *Proc. Roy. Soc.* A, **176,** 180.
Schonland, B. F. J., Hodges, D. B., and Collens, H., 1938, *Proc. Roy. Soc.* A, **166,** 56.
Schonland, B. F. J., Malan, D. J., and Collens, H., 1935, *Proc. Roy. Soc.* A, **152,** 595.
Watson-Watt, R. A., Herd, J. F., and Lutkin, F. E., 1937, *Proc. Roy. Soc.* A, **162,** 267.
Wilson, C. T. R., 1920, *Phil. Trans.* A, **221,** 73.
Wormell, T. W., and Pierce, E. T., 1948, *J. Instn. Elect. Engrs.*, **95** (Part III), 331.

DESCRIPTION OF THE PLATES

The first eight photographs employ the raster technique previously described (§ 2); the starting point upon the raster is indicated by an arrow. The last four, figs. 9–12, contain only one horizontal sweep; in these cases two distinct photographs are shown, one above the other. Line drawings of two of the most complicated atmospherics (figs. 3 and 4) are given in fig. 14 (Plate XV). In these drawings the large arrows indicate the ground wave positions and the small arrows the successive sky waves, according to the best interpretation on the reflection mechanism.

Notes on the characteristics and analysis of the individual photographs follow below. Position references (e.g. 6th trace) refer to traces counted downwards from the top of the figure. The results of analysis by the reflection theory are indicated where applicable, with the appropriate estimates of h (reflection height), D (distance), and the quality of fit ((A), (B), (C) or (D), § 6.2). The location of the storm-centres or the distance of the individual Sferics fix, is also given.

Fig. 1. Oct. 18.00 h

Smooth waveform, recorded at sunset, from Spain, 1300 km distant. Amplitude 0·10 v/m. No satisfactory analysis. 7th, 8th and 9th traces. Four small amplitude (6 mv/m) waveforms. These may be associated with the main atmospheric or may be from other sources.

Fig. 2. Oct. 20.00 h

Irregular high-frequency waveform. Amplitude 0·35 v/m. Storm-centres 400 and 600 km distant. N.W. France.

Fig. 3. Oct. 21.00 h. [cp. fig. 14]

Irregular high-frequency type, followed by a good example of the regular peaked variety, showing a long train of ionospheric reflections. Amplitudes 0·15 v/m and 0·5 v/m respectively. The regular peaked type gives a good analysis (A) with $h=87\pm3$ km and $D=700\pm70$ km. Storm-centre 660–760 km. N.W. France.

Fig. 4. Oct. 19.00 h. [cp. fig. 14]

4th trace. Regular peaked waveform, amplitude 0·4 v/m. Analysis gives $h=85$ km, $D=350$ km, (B).

6th trace. Regular peaked waveform, amplitude 0·25 v/m. Analysis gives $h=86\pm2$ km, $D=370\pm30$ km, (A).

9th trace. Regular peaked waveform, amplitude 0·25 v/m. Analysis gives $h=85$ km, $D=335$ km, (C).

Sferics fix on one of these waveforms in North Sea $D=340\pm40$ km.

Fig. 5. Oct. 20.00 h

Smooth quasi-sinusoidal waveform of amplitude 0·3 v/m asymmetric about the zero due to the presence of a slow tail. No satisfactory analysis. Small amplitude atmospherics on 8th and 9th traces.

Fig. 6. Sept. 14.00 h

A group of typical smooth daytime waveforms. The atmospherics on the 8th and 3rd traces are of amplitudes 0·5 and 0·10 v/m, respectively. Neither gives a satisfactory analysis.

Fig. 7. Jan. 21.30 h

A regular peaked waveform showing clear reflections. Amplitude 0·35 v/m. Superimposed ripple due to long-wave transmission (19·4 kc/s). Storm-centre in Yugoslavia 1500–2200 km distant.

A possible analysis agreeing with the storm position is given by $h=85$ km, $D=1850$ km, (C), but this solution does not fit the first five peaks which depart from their calculated positions by 40–120 μsec.

Fig. 8. Dec. 19.00 h

A complicated example of the regular peaked type. The double positive peaks and the progressive change in the shape of the pulses with increasing order, are often observed upon regular peaked waveforms. Amplitude $>0{\cdot}25$ v/m. Analysis, $h=85$ km, $D=1120$ km, (B). Sferics fix, 930 km, N. Italy.

Fig. 9. Dec. 19.00 h

(*a*) Smooth quasi-sinusoidal waveform. Amplitude 0·11 v/m. No satisfactory analysis. Storms scattered over Atlantic Ocean beyond 2500 km.

(*b*) Smooth quasi-sinusoidal waveform. Amplitude 0·08 v/m. No satisfactory analysis. Position in Atlantic Ocean beyond 4000 km.

Fig. 10. Dec. 19.30 h

(*a*) Smooth quasi-sinusoidal waveform. Amplitude 0·06 v/m. No satisfactory analysis. Storm-centres, Atlantic Ocean, 2400–3000 km.

(*b*) Further example from Atlantic. Quasi-sinusoidal but not so smooth. Amplitude 0·06 v/m. No satisfactory analysis.

Fig. 11. Dec. 11.00 h

(*a*) Typical distant daytime waveform. Amplitude 0·05 v/m. No satisfactory analysis. Sferics fix, S. of Azores, 2900 km distant.

(*b*) Typical distant daytime waveform. Amplitude 0·02 v/m. No satisfactory analysis. Sferics fix, W. Africa, 3 200 km distant.

Fig. 12. (*a*) Feb. 19.30 h

Smooth long oscillatory train. Maximum ripple amplitude, 5 mv/m. Analysis gives $h=80\pm5$ km, $D=5\,000\pm1\,000$ km, (C). Bearings on waveforms of this type, coupled with the analysis, suggest they originate in Equatorial Africa.

(*b*) Nov. 21.00 h

Peaked long oscillatory train. Amplitude of 10th pulse, 50 mv/m. Analysis indicates $h=87\pm3$ km, $D=1600\pm200$ km, (C). Sferic fix, 1400 km, Italy.

XXXVII. *Interferometric Studies of Hardness Test Indentations : Investigations on Tungsten Carbide, Steel, Duraluminium and Tin*

By S. Tolansky and D. G. Nickols
Royal Holloway College, Egham*

[Received January 28, 1952]

Abstract

Studies are made on the surface distortions which occur on metal surfaces which have received hardness test indentations from both diamond pyramid and diamond cone indenters. The surface contours are evaluated by multiple-beam interference methods.

Experiments are reported for indentations over a wide range of hardness, the surfaces studied being those of sintered tungsten carbide, nickel-steel, duraluminium and tin. The distortions found in polycrystalline masses resemble one another but reveal secondary differences.

On single crystallites of tin unsymmetrical patterns are found, with piling-up in two directions and sinking-in in directions at right angles to these. The effects are determined by the directions of the crystal axes.

The concavity and convexity of the sides of the pyramid indentations are discussed in relation to piling-up and sinking-in, and the influence of recovery is indicated.

§ 1. Introduction

In the measurement of metal hardness from the indentations produced, usually by a ball, cone, pyramid or chisel, it has long been recognized that such indentations are accompanied by surface distortions which are described as either ' piling-up ' or ' sinking-in ', and some early approximate mechanical measurements of these effects have been discussed by O'Neill (1934). More recent work is described by Lysaght (1949) and by Crow and Hinsley (1946). O'Neill defines piling-up as the condition in which the perimeter of the surface of contact of the indenter and the specimen is above the original level and it is self-evident from geometry that this leads to a convexity in the superficial outline of a pyramid indentation if distortion, as may be expected, is less at the corners than opposite faces. Conversely sinking-in, with perimeter of contact below the original level, leads to concavity. Lysaght adopts a similar classification, but it should be noted that these descriptions avoid considerations of recovery.

In view of the sensitivity of multiple-beam interference methods it appeared worth while to investigate (using such interference techniques) the surface distortions produced by hardness indentations, since it was clear that the characteristics of such distortions could by such techniques be recognized more exactly and at a far earlier stage than hitherto recorded.

* Communicated by the Authors.

Preliminary experimental results using these techniques have already been reported by Tolansky and Nickols (1949 a, 1949 b) and in these attention has been drawn to the fact that distortions with micro indentations can extend further than was formerly suspected. More recently Hill (1950) has made a mathematical analysis for the hypothetical case of penetration by an infinite wedge and although his results are not strictly comparable with those reported here, they indicate that a trend is to be expected on theoretical grounds similar in nature to those recorded experimentally.

In a group of earlier papers (Tolansky 1942 *et. seq.*) on the theory and application of both monochromatic and white light multiple-beam interference techniques, it has been shown that such procedures offer rapid and powerful methods for the study of surface microtopography, if the surface to be studied is matched against a correctly silvered optical flat. The reflectivity and surface polish of the object under study should preferably be high.

With two high-reflecting surfaces critically illuminated under conditions already previously discussed, it is possible to produce either sharp multiple-beam monochromatic localized fringes which give a precision contour of the surface topography or alternatively white light fringes of equal chromatic order which give in effect a profile of a selected line section.

This paper discusses results obtained by the application of the interference techniques to the study of the distortions produced on metallic surfaces which have received micro indentations, when using (1) a Vickers diamond pyramid with square base and with 136° angle between faces, and (2) a Rockwell diamond cone with included angle 120°.

Four materials of widely differing hardness have been selected for study, namely, sintered tungsten carbide (' Widia H-metal '), a typical standard steel, commercial duraluminium and tin (99·992% purity).

After appropriate surface preparation microindentations were made with hardness indenting machines, applying loads for 15 seconds at room temperature. The indented surfaces were then matched against correctly silvered optical flats.

§ 2. Sintered Tungsten Carbide (' Widia H-Metal ')

The sintered carbide studied, (described as Widia H) is a cobalt-bonded sintered tungsten carbide with hardness exceeding 2 000 on the Vickers scale. It was micro-polished with diamond powder.

Vickers pyramid micro-indentations were made with a load of 500 g. Plate XVI A* shows Fizeau fringes surrounding the ' square ' shaped indentations (magnification is ×400). The fringe quality shows that the surface finish has a high degree of perfection and good reflectivity. The fringes are the arcs of circles in each quadrant around each side of the indent. As the distortion is less than one fringe order, little numerical information can be obtained from this single interferogram, and for fuller evaluation several wavelengths would be needed. In accordance with

* For plates see end of issue.

anticipation the surface distorts to a maximum opposite the centres of the sides of the square indentations, with much less distortion along the prolongations of the diagonals. The metal grain structure size is much less than the dimensions of the indentation.

A more satisfactory picture of the distribution of flowed material is given by the white-light fringes of equal chromatic order taken across a median section, bisecting opposite faces. These are shown in plate XVIB, each fringe being a repeat profile of the one selected section and these demonstrate that the surface rises at an angle of perhaps 3·5° away from the horizontal, from the inner edge of the indentation to a height of only 1 500 A then falls off smoothly. The distortion extends on either side to about twice the diagonal of the indentation. Grain or directional effects do not appear since the material is effectively isotropic.

The outline of the ' square ' indentation exhibits concavity. O'Neill (*loc. cit.*) considers that in general convexity is due to the bulging associated with piling-up and concavity to the hollowing of sinking-in, which follows from geometry if it be assumed that the metal keeps contact with the diamond faces. The degree of concavity is appreciable here, the sides having caved in from the square by an amount easily seen. As near as can be estimated there is a steep drop from the sharp piled-up peak to the indent edge at the undisturbed level. This angle is too small to account for the concavity. The observations establish that concavity can appear associated with small scale piling-up.

This would at first appear to be a geometrical contradiction. A reasonable explanation can however be found in elastic recovery. Initially, during the indentation and whilst the diamond is embedded, any piling-up must lead to convexity. On withdrawal of the pyramid if there is considerable elastic recovery, the face centres, which suffer the greatest strain effect, recover back more than the corner regions and can well overcome the convexity and produce concavity. This can explain then the simultaneous appearance of concavity and slight piling-up. Because of the high interferometric magnification the piling-up in the figures appears appreciable, whilst it is in fact on a very small scale.

Because of the hardness of the material and the danger of damaging the indenting diamond, no deeper indentations were made on the tungsten carbide, and the exploration of the reaction of a surface to regularly increasing loads was carried out with steel and tin.

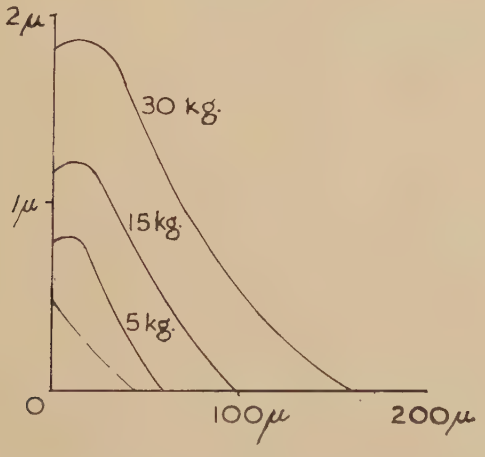
§3. Steel

The steel used (B.S. En 22) contained $3\frac{1}{2}\%$ Ni and was oil quenched, martensitic and with relatively small grain size. It was mechanically polished and this affects the hardness of the layers close to the surface, but possibly not to such an extent as to influence seriously the hardness observations, for the indentations made go in well below the surface. The Vickers hardness number was about 600. Both hardness test and etching showed the absence of any grain orientation and the material can be considered to be effectively isotropic.

Plate XVI C, D, E, showing the fringes (magnification ×120) given respectively by loads of 5, 15 and 30 kg made with a pyramid, present the following features :—

1. The indentation outlines exhibit a slight concavity.
2. The increase in distortion with load is shown.
3. As in the sintered carbide the distortion opposite the corners is less than that opposite the faces, nevertheless with the higher loads it becomes quite appreciable at the corners.
4. The distortions extend beyond the geometrical limits of the indentations, the amount being about twice the diagonal.
5. The fringes curve round on approaching the indentation edge indicating that there is a slight rise from the apparent edge to the maximum of piling-up. As a result of the surface distortion the fringes are not

Fig. 1



Profiles of median sections for steel indents. The dotted curve is the diagonal profile for the 30 kg indent.

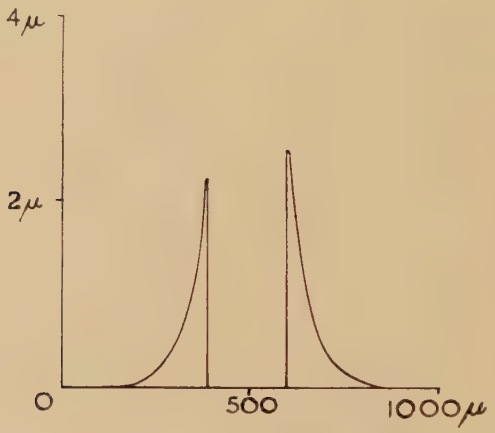
localized at the metal surface, hence it is not possible simultaneously to focus on fringes and on the surface indentation. Plate XVI F is a micrograph of an indentation in the steel. This reveals ' pin-cushion ' distortion (concave sides), a feature shown by all the indents in the steel, and also probably due to elastic recovery.

Fig. 1 shows measured profiles of median sections from edges of the 5, 15, 30 kg indentations. The vertical scale is 100 times the horizontal scale.

There is close resemblance between the profile curves for the different indents. They are almost identically repeated but on increasing scales. In fig. 1 is shown dotted the profile measurement made along the diagonal direction from a corner for a 30 kg indentation. This shows up well the differential effect due to the penetration of the corner for the maximum height in the diagonal direction is only one quarter of that opposite the faces. With loads in the ratio 1 : 3 : 6, the heights of the maxima opposite the faces are in the ratio 1 : 1·5 : 2·3.

O'Neill considers that considerable piling-up is indicative of a low work-hardening capacity. The surface distortion, and consequently the total distortion within the metal, extends appreciably. Any work-hardening effects due to the indent thus extend an appreciable distance beyond the indentation itself. The British Standards Institute Specification No. 427. 1931, defining the conditions applicable to the use of a 136° diamond pyramid hardness test requires that the centre of the impression shall not be less than two and a half times the diagonal of the impression from any edge of the test specimen and from any other impression. It is clear that this accepted safety range of $2\frac{1}{2}\times$ diagonal is just on the limit, for the surface flow around two adjacent indents might well extend slightly beyond this range. When dealing with single crystals the position is complex, and it will be shown that $2\frac{1}{2}\times$ is quite insufficient as a safety margin.

Fig. 2



Profile for cone indent in steel.

§4. Indentations in Steel with a Cone

As the fringe pattern symmetry is determined by the square section of the pyramid, indentations were made in steel with a specially polished Rockwell diamond cone penetrator of cone angle 120°, and as expected symmetrical ring fringes were produced (plate XVII A).

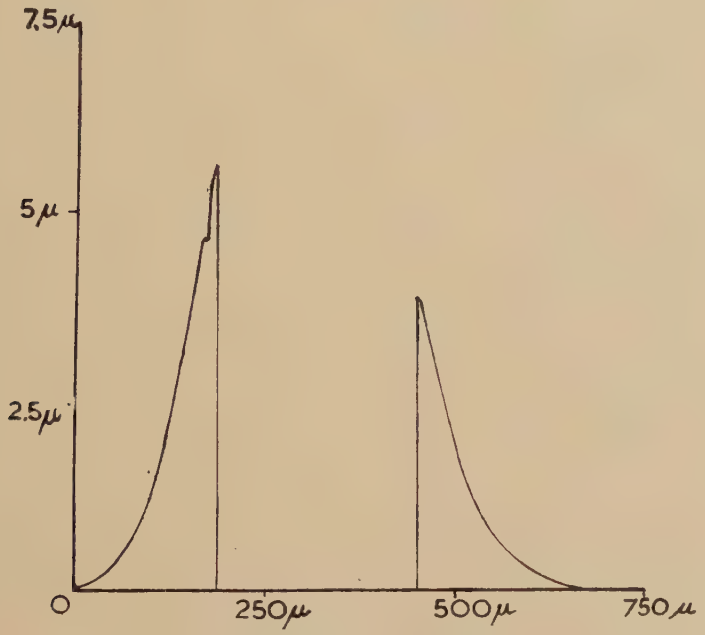
The symmetry and simplicity of the distortion associated with the cone is such that a determination of the volume of extruded material can be readily made. Fig. 2 shows a graphical distribution of a cross-section of the distortion corrected for the wedge angle and since there is rotational symmetry a calculation can be made for comparing the volume of the indent with that extruded. The vertical scale is 250 times the horizontal. Measurement shows that the extruded volume is only some 14% that of the volume of the cone of indentation. In this calculation any possible secondary recovery effects have been neglected since they cannot be evaluated. The fact that the cone does not come to a true point introduces no

serious error into the calculation, a fact established by examining fringes within the indentation itself, these giving a measure of the radius of curvature of the cone tip.

§5. Duraluminium

A specimen of duralumin (hardness somewhat less than 200) was mechanically polished and examined with the Vickers pyramid indenter. The mechanical polish was not of good quality, but some improvement was achieved by electrolytic polishing. Plate XVII B (× 120) shows the indentation with a load of 5 kg Although the distortion resembles that for a steel pattern there are differences. In the steel, the fringes as they approach the edge of the indentation curve round, indicating a drop of some half a light wave. In the duraluminium the fringes run effectively perpendicularly into the edge.

Fig. 3



Profile for indent in duraluminium.

The fringes are irregular and there is evidence of peak doubling in the piled-up region. Fig. 3 gives a smoothed diagram of median profiles which show how the pile-up commences abruptly.

The outline of the indentation is distinctly convex. Thus in the previous two cases where the fringes show either a slight or a steep drop from the pile-up ridge down to the geometric edge there is concavity, but in this case where this drop is absent there is convexity. The indentation with 5 kg has practically the same dimensions as that with 15 kg on steel, but whereas in steel the average pile-up height is about 8 fringes. extending

over a linear range of about twice the side of the indent, in duraluminium the average height is about 18 fringes extending over about $2\frac{1}{2}$ times the side of the indent. The slope in the main part of the pile-up in duraluminium is about 4 times that in the steel, and the extruded volume correspondingly greater. Furthermore the four raised regions of the duraluminium each show evidence of multiple hillocks with secondary valleys between, especially in the near neighbourhood of the outline of the indentation edges, a feature revealed on the left hand side of fig. 3.

The distortion in the direction of the diagonal is relatively small, even less than in the corresponding indent for steel, although the pile-up opposite the face centres is so large.

§6. Tin

The experiments now to be described were made on the very soft metal, tin, and differ from the previous in that the indents were made on individual large single crystallites instead of within a polycrystalline mass. It was considered desirable to use surfaces which had not been subjected to the working which results from mechanical polishing. Such virgin surfaces were produced by casting against a glass flat using the following technique. Tin of 99·992% purity (kindly supplied by the Tin Research Institute) was melted and a small quantity poured onto sheet glass resting on an electric hot-plate, the temperature of the glass being slightly above the solidifying point of the tin. The hot-plate current was cut off and the whole allowed to cool slowly. It was found that the tin conformed closely to the surface of the glass and acquired effectively the smooth polish of the glass. It was removed in disc form by making a diamond scratch on the glass behind the tin disc and on breaking, with a sharp jerky movement, the tin was released. The tin surface was found to exhibit a remarkable degree of 'polish', such as in fact has hitherto not been seen in this laboratory on any metal polished by mechanical means. The reflectivity was of the order of 80% and with such a surface, reflection fringes superior to those given by a mechanically polished metal surface can be formed. The fringe smoothness shows that the metal has taken up the polish of the glass against which it was cast. Furthermore the surface is free from any undesirable cold-work or Beilby layer effects.

Although the surface quality closely approximates to that of the glass, the final tin surface is not as flat, for separate individual large crystallites form with quite distinctive boundaries.

Tin, being soft (Vickers hardness about 5), requires very small loads for the production of quite large indents.

Both pyramid and cone indentations have been made on such individual crystallites. The interference fringes surrounding a pyramid impression produced by a 50 g load is shown in plate XVII C (magnification $\times 160$), and that with 100 g load in plate XVII E (magnification $\times 80$).

A striking asymmetry is evident, yet in fact the asymmetry is even still more marked than it appears, for it is readily established, either by gently

pressing the tin towards the optical flat or by the use of white light fringes of equal chromatic order, that the extended ' wings ' represent *elevations* (piling-up) while the short more compressed pair of contours represent *depressions* (sinking-in). There remains an almost flat region between these, indicated in the plates by the broad fringes and the large areas free from fringes.

Owing to the Feussner surface of fringe localization being at some distance from the metal surface the edges of the indentation are out of focus when the fringes are in focus. Thus to see the shape of the indentation the optical flat is removed and then correct focus can be obtained. The micrograph in plate XVII D ($\times 200$), showing an indentation with 50 g, has an irregular shape, two opposite sides being convex and two concave. This leads to some extent to an optical illusion of distortion, for the corners are not very far from those of a square. On the original of plate XVII D the lengths of the sides measured between corners were respectively 21, 21·5, 20, 19 mm, but the distances between opposite pairs of mid faces were 22 and 18 mm. The mutually opposite curvatures of the sides make the deviation from square outline appear to be in excess of what it is.

The convex sides correspond with the piled-up regions whilst the concave sides are those which have sunk in. This is in accord with O'Neill's principle of simple geometrical considerations, applying if the distorted metal is considered to maintain contact with the diamond and if there is no serious recovery. The striking differences between the piled-up and sunk-in regions are shown in fig. 4. Fig. 4 (*B*) shows a section of the piled-up region from the centre of the face and fig. 4 (*A*) the corresponding sunk-in region.

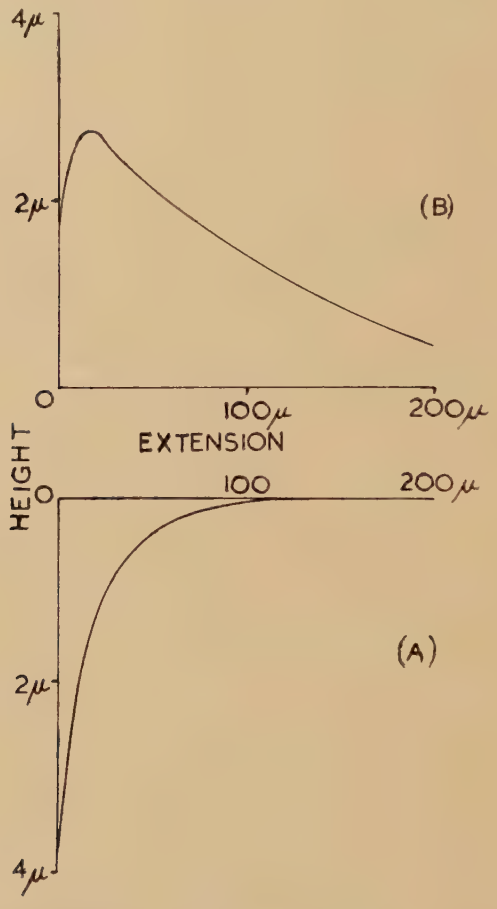
It has been established that this asymmetry in indent and in surface flow is primarily crystallographic and dependent only in a secondary manner on the particular orientation of the square shape of the penetration relative to the crystal axis. Experiments made with the cone indenter reveal asymmetries little different from those given here with the pyramid. The indentation hollow with the cone then appears as an elliptical impression. Across one ' diameter ' the sides of the circle flatten in (sinking-in regions) and perpendicular to this they swell out (piling-up).

It has long been established that crystal faces exhibit oriented hardness differences. O'Neill (1923) first showed this in metals with aluminium crystals, reporting that Brinell hardness values differed in different directions by 13%, leading to non-circular impressions. Schulz and Hannemann (1941) found similar (though somewhat smaller) effects using a pyramid indenter. Meincke (1950) using scratch hardness tests on different faces of single crystals of copper, aluminium and zinc found variations with direction up to as much as 150% in the extreme case. Tests made on the prism faces of zinc and antimony crystals showed differences of 40% between the maximum hardness perpendicular to the *c* axis and the minimum hardness parallel to this.

Thus tin, being tetragonal in structure can be expected to reveal differential hardness effects.

The simultaneous appearance of such marked sinking-in and piling-up is however quite unexpected. In view of the unusually striking patterns which appeared, an x-ray analysis of the orientation of the crystals indented was sought. This was carried out by F. Canepa at Birbeck College who has collaborated in producing the following.

Fig. 4



Profiles of indent in tin. (*B*) Piled-up wings, (*A*) Sunk-in wings.

§7. X-ray Examination of Tin (in conjunction with F. Canepa, formerly at Birkbeck College)

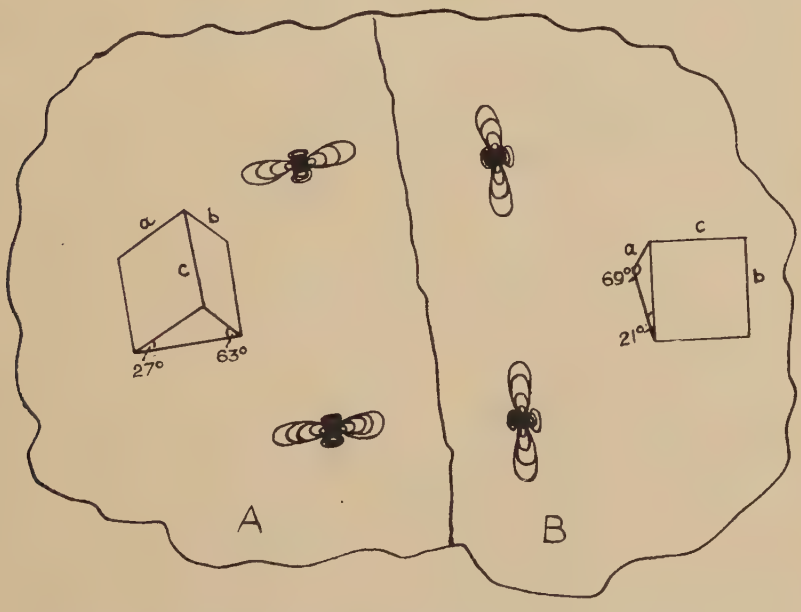
A schematic diagram of the surface of the specimen studied, a disc of radius 1 cm, thickness 7 mm, is shown in fig. 5. This specimen exhibited two large crystal grains, *A* and *B*, and on each were placed four pyramid indents with different loads. In addition two cone indents were placed on *A*. The diagram shows schematically the nature of the surface distortions. On each of the grains the patterns were oriented strictly parallel, proving them to be crystallographic.

The crystallographic orientations of the grains A and B were determined by the method described by Jeffrey (1949) and are shown as a, b, c on each crystallite.

The unit cell of tin is a rectangular parallelopiped with $a=b=5{\cdot}82$ A and $c=3{\cdot}17$ A. Because of the shorter c tetragonal axis an asymmetry was anticipated. In both grains A and B the tetragonal c axis lies at right angles to the pile-up lobes and in both cases is practically on the plane of the surface. In A it is inclined at only 3° to the surface and in B at 5°. The a and b axes in A make angles respectively of 27° and 63° with the surface, and in B of 21° and 69° respectively.

The orientation is rendered schematically. For simplicity c is drawn in the plane of the diagram although in fact it inclines to it at a small angle. The directions of the piled-up lobes are indicated.

Fig. 5



Directional indents and crystal orientation on two crystalites of tin, A and B.

The orientation in grain B is analogous to that in A. The pile-up lobes in both grains are perpendicular to the c axis, which is merely 5° out of the plane.

It may be noted that the a and b axes in grain B make angles of 21° and 69° with the surface, angles which are near to those in grain A. It follows from this that in both grains a (210) plane lies fairly close to the surface. This is fortuitous.

Since in tin many physical properties vary considerably in the different crystallographic directions, differential hardness effects were not unexpected. It is noticeable in figs. XVII C and E that the interferograms

show that the pile-up lobes are not equal on opposite sides and this is in accord with the x-ray result that the a and b axes are differently inclined to the surface. The two small sinking-in lobes on each indentation are much more nearly equal, again in agreement with the x-ray observation that the c axis is nearly in the surface. These observations establish the fact that along the c axis there is sinking-in and in other directions piling-up. One could anticipate that if one met the accidental event of the a–b plane lying in the surface, a symmetrical pattern of the steel type might appear.

§8. Discussion

It has been established by O'Neill, and is confirmed by Heyer (1937), that when dealing with typical polycrystalline metals or alloys, piling-up is evidence of low work hardening capacity and sinking-in of high work hardening capacity. Lysaght points out that cold-work usually results in piling-up with convexity whilst annealed metals tend to produce indents with concave sides. It is clear that in the single tin crystal the two co-existing mechanisms of piling-up and sinking-in are associated with the crystal lattice. There is very little evidence for the existence of slip-planes which if present would have been revealed by the fringes. It has already been established by Tolansky and Holden (1949) that fringes of the type used here are vividly able to reveal slip or twin bands on cast tin crystals.

It is difficult to secure any cast tin surfaces quite free from twin bands over an appreciable area since the act of removal from the glass often produces slight strain which introduces widely spaced twinning lines, always more or less parallel. The slip mechanism is not seriously visibly operating in the case of the indentations studied here.

Bridgeman (1925 and 1933) showed that physical properties of tin crystals, often differ considerably when measured along the c or the a axis. Thus the c/a axis ratios for some physical constants as measured are (1) linear compressibility, $c/a=1/1{\cdot}11$, (2) coefficient of linear expansion, $c/a=2/1$, (3) electrical resistivity, $c/a=1{\cdot}44/1$, (4) magnetic specific susceptibility, $c/a=1/1{\cdot}11$. The present observations illustrate differing behaviour in the c and a directions, with regard to surface flow.

The observations on the sintered carbide, steel, duraluminium and tin suggest that the dual relations ' concavity→sinking-in ' and ' convexity→ piling-up ' are not as simple as had been formerly supposed. The carbide indents are concave yet there is clear piling-up in addition, and the same feature repeats in all the steel interferograms.

It might well be that any concavity associated with a definite large piling-up is always due to elastic recovery and is not due to a simple geometrical conformation of the metal contacting the diamond, for such a conformation alone must produce convexity with piling-up.

The interferograms taken with tin reveal one feature to which attention should be paid in connection with microhardness testing of single crystallites, a technique which is becoming widespread. For in the specially

favoured direction for piling-up the distortion in tin can extend to beyond four times the length of the diagonal. It follows that neighbouring indentations should be separated by at least *eight* times the diagonal if they are not to suffer disturbance. This is to be compared with the $2\frac{1}{2}$ times value adopted in British Standards Institute specification for polycrystalline materials. Microhardness tests are now often being made on individual crystallites and alloys, and when placing indents on such crystal grains, the accepted safe separation of $2\frac{1}{2}$ times the diameter will be insufficient if the crystal has differential effects corresponding to those on tin. As the directional hardness effects of most crystals are as yet unknown, prudence demands that in general the separation should well exceed the $2\frac{1}{2}$ times value, and the same reasoning applies to the distance of the indent from crystal edges.

Acknowledgments

This investigation has been supported with grants made by B.I.S.R.A. and D.S.I.R. for the purchase of apparatus and with a bursary from B.I.S.R.A. The sintered carbide was polished by Industrial Distributors (Sales) Limited through the courtesy of P. Grodzinski, who also prepared the Rockwell diamond cone. Dr. Hedges and Dr. Cuthbertson of the Tin Research Institute have assisted materially in discussion and also by supplying high purity tin.

Particular thanks are due to Dr. Canepa, formerly of Birkbeck College, for generous co-operation in independently carrying out the x-ray examination of the tin.

References

Bridgeman, P. W., 1925, *Proc. Amer. Acad. Arts. Sci.*, **60,** 305; 1933, *Ibid.*, **68,** 95.

Crow, T. B., and Hinsley, J. H., 1946, *J. Inst. Metals*, **72,** 461.

Heyer, R. A., 1937, *Proc. Am. Soc.*, Testing materials, **37,** 11.

Hill, R., 1950, *Phil Mag.*, **41,** 745.

Jeffrey, J. W., 1949, *Acta Cryst.*, **2,** 15.

Lysaght, V. E., 1949, *Indentation Hardness Testing.* Remhold.

Meincke, H., 1950, *Z. S. Metallkunde*, **41,** 344; *Industrial Diamond Review* 1951, **11,** 1.

O'Neill, H., 1923, *J. Inst. Metals.*, **30,** 289; 1934, *The Hardness of Metals and its Measurement.* Sherwood Press.

Schulz and Hannemann, 1941, *Z. S. Metallkunde*, **34,** 124.

Tolansky, S., and Holden, J., 1949, *Nature, Lond.*, **164,** 754.

Tolansky, S., and Nickols, D. G., 1949 a, *Nature, Lond.*, **164,** 113; 1949 b, *Ibid.*, **164,** 840.

XXXVIII. *Twin Formation in Cadmium*

By N. Thompson and D. J. Millard
H. H. Wills Physical Laboratory, University of Bristol*

[Received January 30, 1952]

Summary

Experiments on the formation of twins in single crystal wires of cadmium under tension demonstrate the essentially discontinuous nature of the process of twin growth, and produce evidence in favour of a critical resolved shear stress law for twinning. Twinning can contribute to creep at room temperature but not at 90° A. A dislocation mechanism for the growth of a twin is discussed in some detail.

§1. Introduction

It has been known for many years that deformation twins can readily be produced in hexagonal metals. Most of the experimental work has been done on magnesium, cadmium and zinc, these being technologically the most important examples of the hexagonal structure. But surprisingly little attention has been paid to the detailed mechanism of the twinning process, in contrast to the voluminous literature on deformation by gliding. The crystallography of twin formation has been established beyond reasonable doubt, but little is known of the forces required to produce the change, and of the atomic movements involved.

In this paper we give first of all a brief review of previous work (§ 2). This is followed by a description of some experiments on twinning in single crystals of cadmium (§ 3). Finally, we discuss the conditions under which twin formation may occur and offer some suggestions, largely speculative, concerning the mechanism by which the changes might be brought about (§ 4).

It should be noted that although the experimental work was done on cadmium there is every reason to suppose that the results for zinc would be qualitatively similar, and that the same would apply to magnesium also with appropriate changes of sign.†

§2. Previous Work

It was established by Mathewson and Phillips (1927), and subsequently confirmed by other workers, that the twinning plane in cadmium is always one of the six $\{10\bar{1}2\}$ type planes. The composition plane of the twinning is also approximately $\{10\bar{1}2\}$. In structures of higher symmetry

* Communicated by the Authors.

† The possibility of twin formation of the type here considered appears to depend on the fact that the axial ratio of the hexagonal lattice (c/a) is not equal to $\sqrt{3}$ (see Barrett 1948). For both Zn and Cd, $c/a > \sqrt{3}$ while for Mg and all other metals $c/a < \sqrt{3}$. Many of the formulae describing the process change sign when c/a passes through the value $\sqrt{3}$. Hence the above qualification.

(such as the body centred cubic lattice) it is possible to imagine a region of crystal transformed into its twin by a homogeneous shearing motion parallel to the twin plane ; the twinning direction is the direction in the twinning plane, in which this shear takes place. In the hexagonal metals this is not possible, as was shown by Kolesnikov (1933) and discussed by Barrett (1949) ; only half of the atoms involved can be moved from the matrix to the twin position by a simple homogeneous shear. The remainder must, in addition, move in various directions which are however, all contained by the $(1\bar{2}10)$ plane ; this is the plane of shear and is that plane normal to the twinning plane, which contains the twinning direction. Even so, the macroscopic effect is one of simple shear and it is useful to speak of the twinning direction, and of the twinning shear. For a hexagonal metal, the latter quantity depends on the axial ratio c/a and is given (Schmidt and Boas 1935) by

$$S=\{(c/a)^2-3\}/\sqrt{3}\,.\,c/a.$$

Thus for cadmium and zinc, for which $c/a>\sqrt{3}$, S is positive while for all other metals it is negative.

It is clearly a necessary condition for twinning that the shear stress resolved on the twinning plane and in the twinning direction, shall have the same sense as the twinning shear. By analogy with the process of slip one is tempted to take the obvious next step and suggest that a sufficient condition for twinning is that the resolved shear stress in the twinning direction shall reach a certain critical value characteristic of the crystal. Such an assumption has previously been made by several workers. If one is prepared to admit it, some results of Boas and Schmidt (1929) enable one to make a (rather rough) estimate of the value of the critical stress for cadmium. They report that twinning commenced in a tensile test when basal glide had caused the (0001) plane to rotate so that it made an angle of 9° or less with the specimen axis and that the stress to cause twinning increases with the amount of preceding basal glide. From the published results one can deduce that in the limiting case of little or no such glide (i.e. when the crystal orientation was initially favourable for twinning) the tensile stress at the onset of twinning was of the order of 1 000 g/mm^2. Under these conditions the twinning plane would be at about 45° to the axis and thus the resolved shear stress would be about 500 g/mm^2.

This same assumption of the existence of critical resolved shear stress for twinning is taken by Davidenkov, Kolesnikov and Federov (1933) as being ' very plausible ' and used by them to predict which of the possible twin planes would in fact be active in their experiments. These were made with single crystals of zinc tested in tension, and they report that the twin planes observed agreed with those predicted on the above basis. Nevertheless, and on the basis of rather inadequate evidence, they conclude that the proper criterion for the onset of twinning is something more complex than the above. Their experiments were few in number, however, and it would be easy to attach too much importance to this

opinion. If such a critical stress exists, then the published data of these workers provide four determinations of it for zinc at room temperature, which range from about 390 to 610 g/mm^2. Similarly Miller, also working with single crystals of zinc, quotes eight observations of this resolved shear stress at the instant when twinning was first detected (by the 'characteristic snapping sound'). These appear to be spread over a wide range—from 300 to 650 g/mm^2—but if we neglect the solitary observation for the crystal whose base plane was almost parallel to the specimen axis the range is much reduced (300–450).

Bakarian and Mathewson (1943) experimenting on a magnesium single crystal found that it formed twins readily when compressed normal to the c axis as would be expected. Numerical values for the stresses at the onset of twinning are not given but a rule is formulated for determining which of the possible twin planes is in fact operative. This is approximately equivalent to the statement that the favoured planes are that pair for which the resolved shear stress under discussion is the greatest. The only other experimental work known to the authors which bears on this question is that of Gough and Cox (1929 and 1930). These workers tested single crystals of zinc in torsional fatigue and showed that where twin bands appeared on the surface the twinning plane was the same as that mentioned above. However, they were unable to establish, as a result of their observations, any criterion which would predict correctly which of the six possible twin planes would be active in any region of the specimen. In particular, the occurrence of twins was not closely correlated with the maximum value of shear stress across the twin plane and in the twinning direction. It may be that the more complex stress system involved, and the fact that there were considerable stress gradients present are in some way responsible for this failure of a criterion which appears to be reasonably satisfactory in other experiments.

§3. Experiments on Twinning in Cadmium

In this section we describe some experiments that we have made which have a bearing on these questions. They concern the conditions under which twinning occurs in single crystals of cadmium under tension. The material used was 'Avonmouth' cadmium* 99·98% pure, in which the principal impurities were Zn 0·01%, Pb 0·007%. Wires about 2 mm diameter were grown into single crystals in air, by the method of Andrade and Roscoe (1937). The orientation of the crystals was determined by examining an etched sample on a 2-circle optical goniometer (Barrett 1943).

Crystals a few centimetres long were suspended vertically and loaded by loops soldered to their ends. Two loading devices were used. The first consisted of a weight hanging from the bottom of the crystal, and

* Kindly supplied, and analysed, by the National Smelting Co. Ltd.

partly immersed in a tank of water. As the water was slowly drained away, the applied load increased very smoothly. The second permitted sudden increments of load to be applied ; it consisted of a slug of iron, interposed between the crystal and the weight just mentioned, which was attracted downwards by a fixed solenoid.

One of the objects of the experiments was to investigate the sudden and discontinuous nature of the twinning process, which has frequently been remarked. To this end, the twinning was detected by suspending the specimen from a short cross-bar which rested on a group of piezo-electric (Rochelle salt) crystals. When a portion of the cadmium crystal was transformed into the twin orientation, an impulse travelled along it and gave rise to a signal from the piezo-electric pick-up. This was amplified, displayed on a cathode ray oscillograph, and recorded on a moving film camera. In addition, an optical extensometer was used to record the changes in length of the specimen : this too gave a continuous photographic record. The magnification of the optical lever was about 800.*

The observations made with this apparatus fall conveniently into three groups :

(i) When the load on a single crystal specimen was gradually increased, there came a time when the piezo-electric pick-up began to record a number of small irregularly spaced jerks. Photographs taken with a high-speed drum camera in conjunction with the C.R.O. (writing speed $\approx 10^3$ cm/sec) showed that an elementary jerk consisted of an irregular, damped train of waves lasting 1/100 sec, or rather more, in which the predominant period was of the order of magnitude of 10^{-4} sec. The wave form often varied considerably from one record to another ; the only consistently reproducible feature was at the commencement, where it showed an accelerating rise to a sharp peak in about 150 microseconds, followed by an even more sudden fall. The sign of this initial peak was always the same, and corresponded to a transient reduction of the steady load on the Rochelle salt. It is conjectured that the details of the recorded impulse correspond to shock excitation of the various modes of vibration of the rather complex mechanical structure used to support and load the specimen ; as such they are of no interest. It would appear likely that the duration of the actual elementary process of twin formation is very short—perhaps about 10^{-4} sec, or even less.

The jerks were usually recorded on a slower moving film camera (writing speed=35 cm/sec) and, as no interest attached to their structure, they were rectified and partially smoothed before being fed into the C.R.O. This simplified the interpretation of the records, each jerk appearing as a peak with a steep front, decaying with a time constant of about 1/100 sec. It was first necessary to show that these jerks were

* The experiments formed part of the work of one of us (D.J.M.) for the degree of Ph.D. at Bristol University and a detailed description of the apparatus has been given in the Ph.D. Thesis (1950).

in fact correlated with the formation of twinned regions. It was observed that some specimens showed the effect more readily than others and that the differences depended on the orientation of the *c*-axis of the cadmium crystal relative to the specimen axis. Favourable orientations were when the angle between the specimen axis and the basal plane (usually denoted by χ) was small, preferably less than about 6°. Under these conditions the resolved shear stress on the slip plane (the basal plane 0001) would be small, while one of the six possible twinning planes would make an angle of the order of 45° with the wire axis, and consequently the resolved shear stress on it would be large. Thus the conditions might be expected to favour twinning as against slip.

The question seemed to be settled beyond reasonable doubt by one group of observations. A single crystal was loaded gradually until jerks were being recorded: the load was then held constant for some minutes and finally removed. The recording extensometer showed no appreciable elongation, but the oscillograph records showed four well-marked jerks, all within an interval of 4 sec. Microscopic examination of the surface of the crystal at this stage showed the presence of four straight narrow bands running obliquely right across it, two at each end; no such markings had been visible at the commencement of the experiment. With the microscope available, the widths of these bands could only be estimated approximately and were of the order of 5 microns. The calculated extension due to the formation of twin bands of this width is too small to have been detected by the extensometer.

This visual examination was done without removing the specimen from the apparatus, and when it was complete the load was again applied slowly, until more jerks were recorded. This time they were so numerous that the amplifier system was overloaded and no quantitative conclusions can be drawn from the oscillograph records. Re-examination under the microscope showed that no new bands had been formed but that all four of the old ones had become wider. The gap between the two members of each pair (about 20 microns) had, as near as could be judged, remained constant, i.e. all four bands had grown outward. At this stage the specimen was removed from the apparatus, etched, and examined on the goniometer. Stereographic plots of reflections from both matrix and the four bands showed that these bands were indeed twins, formed on the two twinning planes over which the resolved shear stress had been greatest. The formation of this much twinned material in the orientations observed was calculated to give an extension of the specimen of $2{\cdot}18\ 10^{-3}$ cm. This should have produced a deflection of 1·73 cm on the recording extensometer; actually, a movement of 1·65 cm was observed, an excellent agreement.

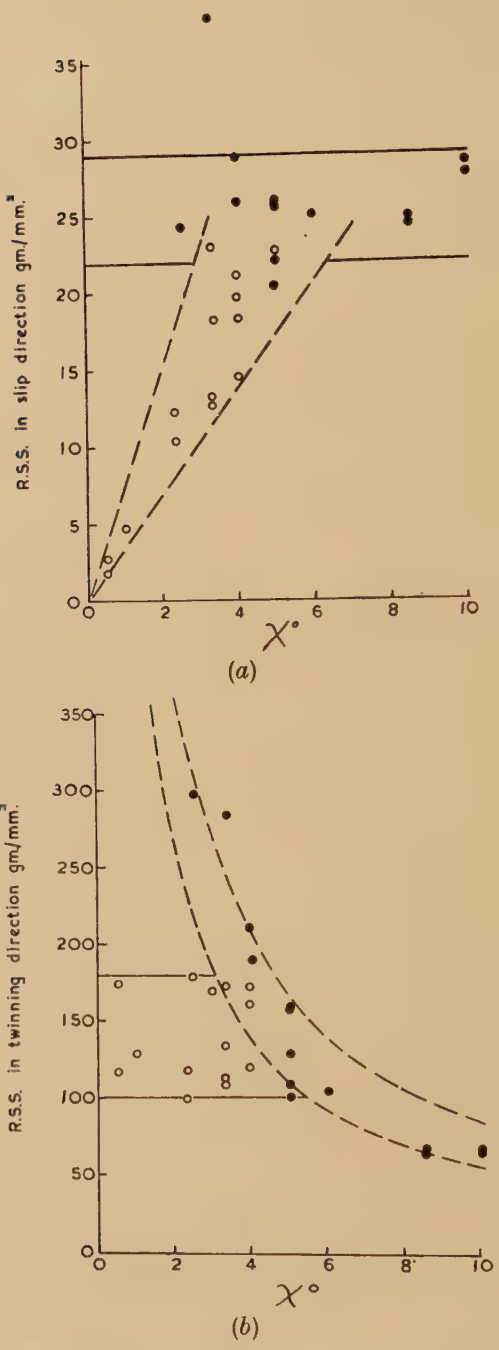
From these observations we may conclude (*a*) that the jerks recorded are in fact caused by the formation of twin bands, (*b*) that the process of formation and growth of a band is essentially a discontinuous one and (*c*) that the elementary process is probably very rapid.

(ii) The specimen used for the observations just described was not of uniform cross-section, and so was not suitable for the measurement of the stress necessary to cause twinning. For this purpose a number of crystals were selected whose orientations were such that χ was less than 10°. Their sectional area was determined by weighing a measured length and they were then mounted and tested in the way already described, the tensile stress being increased at a constant rate of about 1 g/mm^2/sec. Synchronized records were made using both the extensometer and the jerk recorder. From the latter one could easily deduce the stress at which jerks first appeared. The former frequently showed a sharp knee in the curve, indicating a sudden increase in the rate of extension. The stress at which this occurred was also noted and found not to coincide with the beginning of the jerks.

The crystals had been grown in contact with air, and under these conditions it is known (Cottrell and Gibbons 1948) that cadmium shows a well-defined yield point at the lowest stress at which deformation by slip will occur. It was conjectured that the sharp knee in the extensometer records was caused by such yielding while the onset of jerks denoted the start of twinning. To test this hypothesis the two critical stresses were resolved, firstly on the slip plane and in the slip direction, and secondly on the twin plane and in the twinning direction. (The active twinning plane was assumed to be that one of the six on which this resolved shear stress was the greatest. On the basis of this assumption one can calculate the orientation of the twinned material and compare the result with the orientation as determined using the goniometer. This was done on about half the crystals measured and the agreement was satisfactory. It may be noted in passing that this circumstance in itself speaks in favour of the existence of a resolved shear stress criterion for twinning.)

The resulting critical stresses are plotted against the angle χ in figs. 1 (*a*) and (*b*). In the former, the stresses have been resolved on the slip plane and it will be seen that the full circles (corresponding to the knee in the extensometer record) do in fact cluster about a horizontal line, as would be expected, with a mean value of 26 g/mm : the open circles, which we claim denote the onset of twinning, do not. In fig. 1 (*b*) the same critical stresses have been resolved on the twinning plane and this time it is the open circles which cluster about a horizontal line, while the others do not. The mean value of the critical resolved shear stress for twinning, found in this manner, is 138 g/mm^2, but the spread about this mean value is considerable ($\pm$40 g/mm^2) and is a good deal larger than the estimated maximum experimental error (about $\pm$10 g/mm^2). Nor do the points show any systematic variation if plotted against any of the other angular parameters which might reasonably be involved. In all, 26 specimens were tested in this way, and in 14 of them twinning occurred before slip. In those instances where slip occurred first the critical stress for twinning is not recorded, since the twinning might be taking place in a lattice already deformed by slip.

Fig. 1



Resolved shear stresses (*a*) on the slip plane and in the slip direction, (*b*) on the twinning plane in the twinning direction. Full circles—yield point : open circles—onset of twinning. Room temperature. χ denotes angle between basal plane and specimen axis.

A similar set of observations was made with the cadmium crystal immersed in liquid oxygen. 19 specimens were tested, 10 of which twinned before slipping. The results are shown in the same manner as before in figs. 2 (*a*) and 2 (*b*). The values of the critical resolved shear stresses are 87 g/mm^2 for slip and 295 g/mm^2 for twinning—both considerably greater than at room temperature.

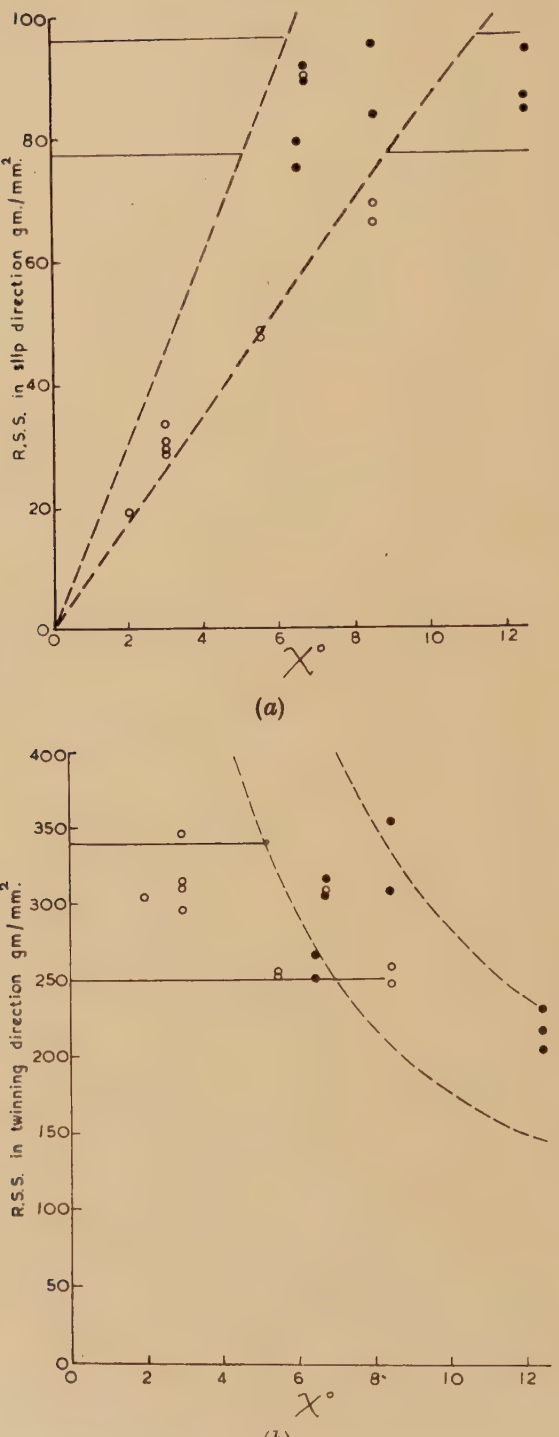
The values of the critical resolved shear stress for slip are of the same order as those usually found for cadmium, albeit the variation with temperature is greater than that recorded for example, by Schmidt and Boas. But the range of orientations of the crystal in which twinning occurs before slip is so small, and the spread of the results referring to twinning is so great, that little more can be said than that they are not inconsistent with the existence of a critical resolved shear stress criterion for the onset of twinning. This is plausible, and is supported by other evidence already mentioned. If such a criterion does exist, the value of the critical stress given by those measurements is about 138 g/mm^2 at room temperature and 295 g/mm^2 at 90° A.

(iii) The third group of experiments was concerned with the effect of applying a sudden load increment to a specimen in which a small amount of twinning had already taken place. The increment was made by using the electro-magnetic device previously mentioned. The load was held constant at a value slightly above the critical value for twinning until all activity had ceased. The increment was then applied and held constant. In all cases, at room temperature, it was some tens of seconds after the increment had been added, before the resultant twinning ceased. These jerks appeared in irregular groups, becoming less violent and less frequent until they finally stopped altogether.

It was desired to see if there was any quantitative correlation between the recorded creep extension, and the amount of twinning as indicated by the recorded jerks. It was assumed that the height of a peak in the jerk record was some measure of the amount of crystal converted into twin on an arbitrary scale. The cumulative total of the peak heights, from the moment of application of the load increment, could be measured on the jerk record, and was called the ' amount of twinning '. Figure 3 shows the result of one such experiment. The two curves show, respectively, the amount of twinning as just defined, and the total extension calculated from the extensometer record. Both quantities are plotted against time and the relative vertical scales have been chosen so that the two curves coincide at the point *X*. Two successive increments of load were made on the same specimen and it will be seen that on both occasions there is a general agreement between the shapes of the two curves representing extension and amount of twinning.

The same result was obtained from a second similar experiment, while a third gave the results shown in fig. 4. The agreement is satisfactory for the first increment, but not for the second. The data for all three are summarized in table 1. From this it will be seen that the stresses involved are higher in the third experiment than in the other

Fig. 2

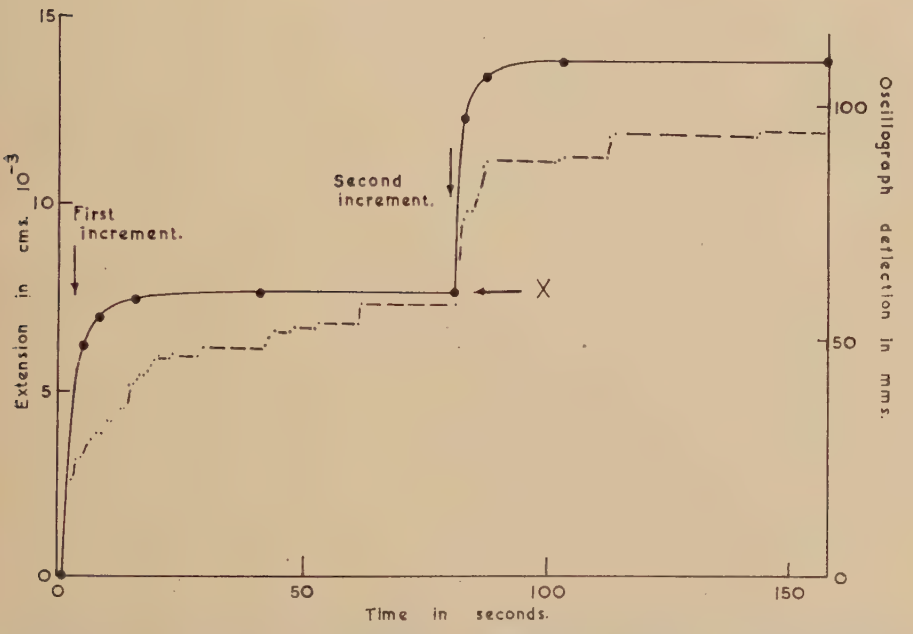


Resolved shear stresses (*a*) on the slip plane and in the slip direction, (*b*) on the twinning plane in the twinning direction. Full circles—yield point : open circles—onset of twinning. Temperature 90° K. χ denotes angle between basal plane and specimen axis.

two, and that the second increment in particular was rather large and probably caused the resolved shear stress on the base plane to exceed the yield stress. Thus the excess extension observed can with some confidence be attributed to slip and not to twinning.

When any considerable amount of twinning has taken place, the possibility of slip on the basal plane of the twinned regions must be considered. If the crystal orientation is known, the value of the appropriate resolved shear stress can be calculated. The values are given in table 1 and are seen to be considerably above the critical value for slip in an unconstrained crystal. Whether or not they would be

Fig. 3



Correlation between creep and twinning: Specimen No. 89b. Room temperature. Full line—creep extension: dashed line—amount of twinning on arbitrary scale, see text.

sufficient to cause slip in a thin lamina of twin bounded on both faces by matrix is more doubtful, but the possibility of this happening will increase as the thickness of the lamina increases. Thus it may be that some of the observed extension is due only indirectly to the twinning; certainly on some occasions slip lines parallel to the basal planes of twinned regions could be observed under the microscope.

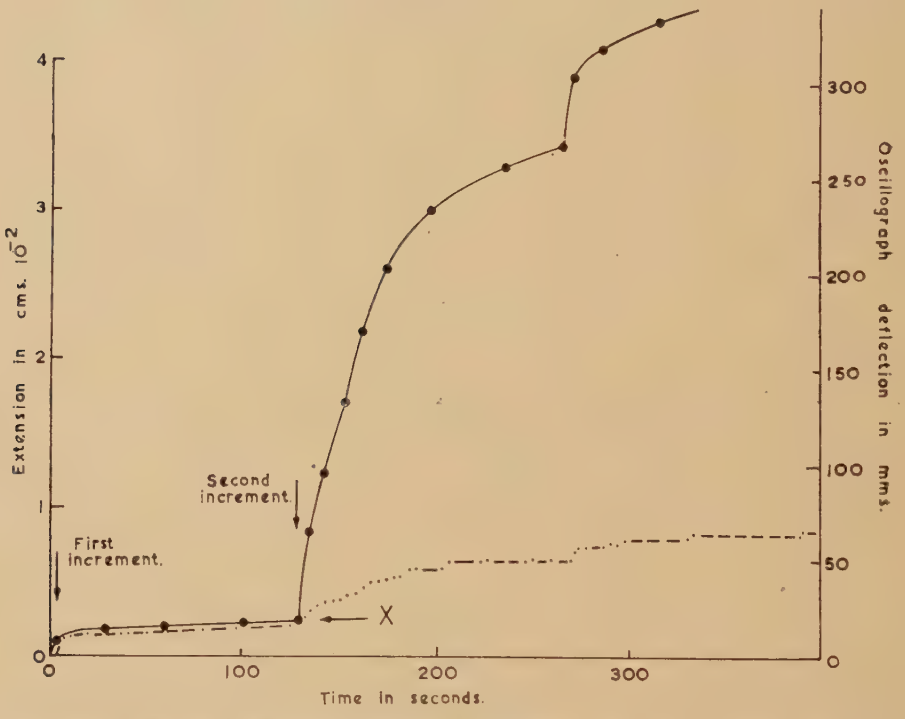
Experiments similar to the above were made on four specimens at the temperature of liquid oxygen, two successive load increments being applied to each. The jerk records showed that, on every occasion, any twinning which took place did so immediately* the load increment was

* i.e. within an interval of less than 1/5 sec, during which time the jerk record was perturbed by pick-up from the electromagnet.

applied. There was no ' creep by twinning ' such as was observed at room temperature, although at the higher stresses the extensometer records sometimes showed a small amount of creep, attributable to slip.

It will thus be seen that both at room temperature and at low temperatures the processes of twin formation and growth are essentially discontinuous. From an inspection of the jerk records made with slowly

Fig. 4



Correlation between creep and twinning : Specimen No. 80c. Room temperture. Full line—creep extension : dashed line—amount of twinning on arbitrary scale, see text.

Table 1

Specimen Number	Resolved shear stress in : Slip direction	Twinning direction	Slip direction in twin	Orientation χ	λ
89b	11·7/13·8/15·8	111/131/150	117/139/158	2·3°	3·3°
73b	16·6/19·2/24·0	123/142/178	104/120/150	3·3°	8°
80c	24·5/29·0/35·7	160/190/234	101/120/147	4·0°	5°

The three numbers on each group denote respectively the initial stress and the stress after the first and second increments. Stresses in g/mm^2. χ= angle between specimen axis and basal plane. λ=angle between specimen axis and slip direction.

increasing loads, one can say that the elementary process is of the same order of magnitude at both temperatures. The difference is that at low temperatures all the twinning takes place within a very short time of a load being applied whereas at room temperature it may proceed in irregular bursts for some tens of seconds afterwards.

(iv) The general picture of the growth of a twin that one obtains from these experiments is as follows: an existing configuration of matrix and twin is stable, until the applied stress reaches a critical value, which is greater the lower the temperature. The exact criterion is possibly that the resolved shear stress on the twinning plane and in the twinning direction should attain a critical value, but this, although plausible, is not established with certainty. If it is correct, then the reproducibility of the experimental results is not good. This in turn might indicate that some accidental singularity near the existing twin matrix interface determines the precise value of the average stress—which is all that is measured—at which the twin begins to grow. Be that as it may, when the critical value is reached the twin-matrix interface moves forward a short distance into the matrix and then again comes to rest. The time occupied by this movement appears to be quite short, probably not more than 10^{-4} sec and possibly less. The distance advanced is not always the same, but is usually of the order of magnitude 10^{-4} cm.* Larger movements are built up of a number of such steps in rapid succession. At room temperature, and under constant stress, this process can be repeated in an irregular sequence but with diminishing frequency for some seconds. In general terms this suggests that some local obstacle holds up the advance of the twin matrix interface temporarily, and that under the combined action of the applied stress and thermal fluctuations the obstacle is eventually overcome, and the interface moves forward until it meets the next obstacle. This is in agreement with the observed fact that such a process does not occur at 90° A. It remains only to explain why this forward motion of the interface does eventually stop altogether, and yet can be started up again by a further increase in the applied stress.

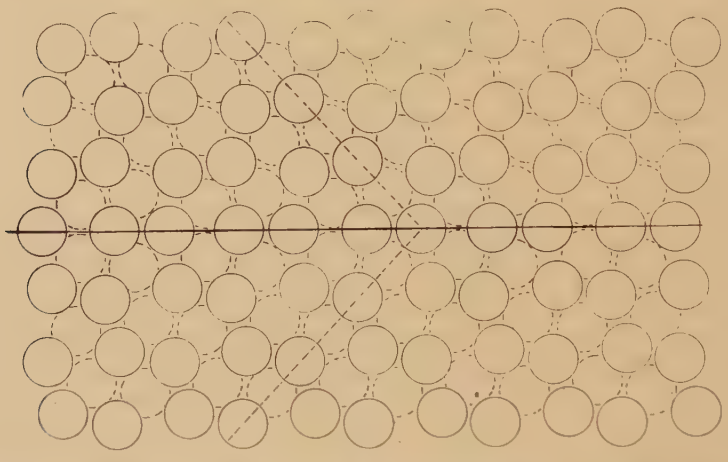
§4. The Mechanism of Twinning

As already mentioned, the crystallography of twinning in cadmium is well established (Mathewson and Phillips 1927, 1928) and one or two authors have speculated on the atomic movements involved, and on the arrangement of the atoms near a twin boundary (Mathewson and Phillips 1927, Mathewson 1928, Gough and Cox 1929). The most recent discussion is that of Barrett (1949) in which he points out that the composition plane of the twinning, e.g. $(10\bar{1}2)$ is 'corrugated'. If we denote successive close-packed (0001) layers of the hexagonal lattice as A and B, then successive $[1\bar{2}10]$ rows of atoms lying in the composition

* It should perhaps be remarked that these figures refer to observations made on wires about 2 mm diameter. A few measurements were made on thinner wires, with essentially the same results, but the range of size was not great, and the possibility of a size effect is not excluded.

plane, belong to A and B planes in turn, and lie alternately slightly above and slightly below the mean level of the $(10\bar{1}2)$ plane. The plane is thus in a sense a double layer, and Barrett refers to the two parts as $(10\bar{1}2\ a)$ and $(10\bar{1}2\ b)$. The same will clearly be true of the lattice of the twinned portion. The question now arises as to how the two lattices are to be fitted together. Barrett suggests four possibilities: (i) the common $(10\bar{1}2)$ plane is a type in both twin and matrix, (ii) it is b type in both twin and matrix, (iii) it is a in twin and b in matrix, (iv) it is b in twin and a in matrix. If the appropriate diagrams are drawn out to scale it will be found that one of the first two will lead to a zone near the twin boundary where the density is appreciably less than in normal lattice, while the other gives a corresponding zone of increased density. Both situations can be relieved by permitting the two halves of the diagram

Fig. 5



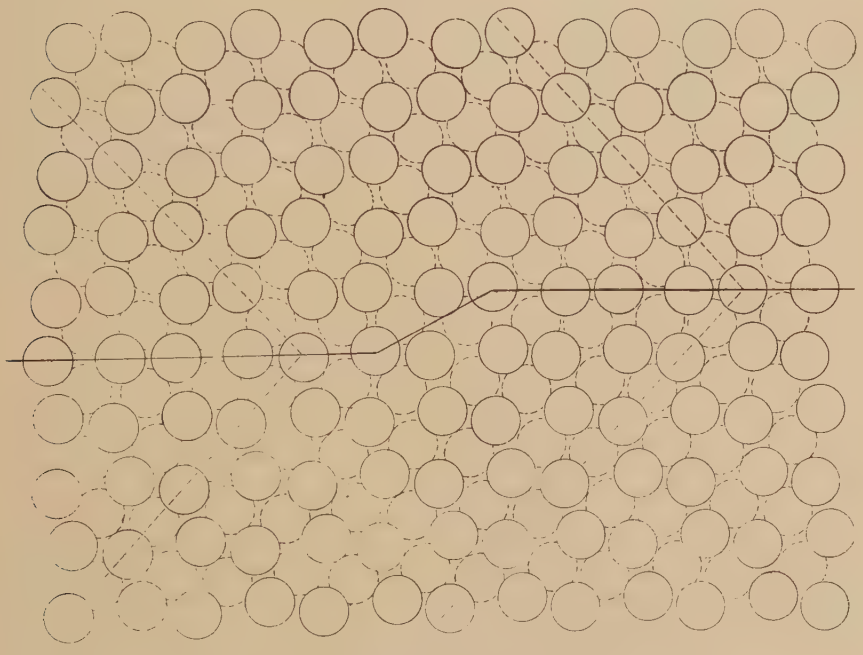
Suggested arrangement of atoms near a twin boundary. Oblique dashed lines show traces of basal planes.

to move towards (or away from) each other. If this is done, and if the necessary symmetry of the boundary is established by one or two minor re-adjustments of atomic position near to it, the two diagrams become identical. They are identical also with the two possibilities (iii) and (iv) after these, too, have been subject to minor adjustments to establish the symmetry and to relieve the more obvious misfits. Thus all four possibilities lead to the same final picture of the twin boundary. The arrangement thus established is shown in fig. 5 where the atoms are represented by spheres that would be close-packed on a basal plane.

It is well known that the composition surface between matrix and twin is only exceptionally plane: twin inclusions are usually lenticular, with curved boundary surfaces coinciding only approximately with $\{10\bar{1}2\}$. We are thus led to consider the possibility of a step in a twin boundary. Figure 6 shows a suggested atomic arrangement near such a step drawn in a similar intuitive manner to that in which the boundary arrangement

itself was chosen. It will be seen that the disturbance of the lattice is nowhere very violent, i.e. the energy of such a configuration will not be impossibly high. Moreover, only minor movements of a comparatively few atoms situated near the step, are required to allow the configuration to move along the twin boundary. If it were to do this, the effect would clearly be to transform a layer of the lattice from matrix to twin (or vice versa, depending on the direction of movement). The singularity is, in fact, a twinning dislocation as already discussed in general terms by Frank (1951) and Frank and van der Merwe (1949).

Fig. 6



A (double) step in a twin boundary.

The configuration shown in fig. 6 is actually a double twinning dislocation, the minimum step height being one-half of that shown. This arises because the atoms shown by dotted lines, which lie above (or below) the plane of the diagram are in every way equivalent to those shown in full lines, and a possible position of the composition plane will be half-way between the two shown. Since the two components of such a double dislocation will repel one another, the precise situation shown in fig. 6 will not be found in practice.

It is suggested that the growth of a twin takes place as the result of repeated passages of such twinning dislocations across the twin boundary, under the action of the applied stress. If this is correct, then at any one time nearly all relative atomic movement is confined to a comparatively small number of atoms in the immediate neighbourhood of the instantaneous position of the twinning dislocation(s) involved.

It is clear from the above argument that the existence of non-planar composition surfaces requires the existence also of twinning dislocations or some equivalent singularities. As in the case of the more familiar dislocations we may also argue that either (i) the transformation of all the atoms in a complete atomic layer from matrix to twin configuration takes place simultaneously—which is, on the face of it, somewhat improbable or (ii) it does not—in which case the boundary between the transformed and untransformed part constitutes a loop of twinning dislocation.

The essential similarity between an ordinary dislocation and the twinning dislocation can be seen as follows. Let $ABCD$ (fig. 7 (a)) represent a portion of a crystal which can twin on planes normal to the paper and parallel to AB. Make a hole normal to the paper at O and a cut in the crystal along OX. Let OY, OZ be traces of two possible composition planes separated by the minimum repeat distance (x). Now convert the two portions YX, ZX to twin configuration as in fig. 7 (b).

Fig. 7

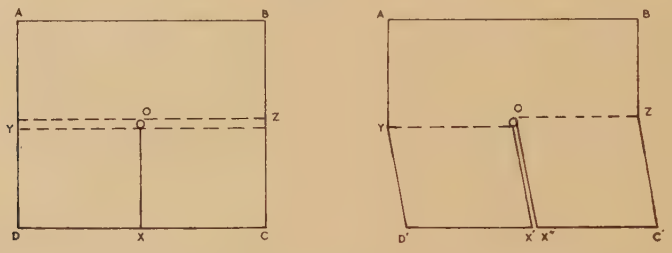


Diagram to illustrate the nature of a twinning dislocation.

This can be done by a homogeneous shear without introducing any internal stresses, but the two faces OX', OX'' will now be separated by a distance (x . s) measured parallel to AB, where s is the twinning shear. The two faces—themselves undistorted—can now be re-united (apart from the singularity at O) and the twinning dislocation and its associated stress field will have been introduced into the crystal. It is thus clear that the stress field around such a twinning dislocation will be of the same kind as that around an ordinary edge dislocation, and that the magnitude of the Burgers vector (b) is the quantity (x . s). This need not be a lattice vector, and is in fact usually a fraction thereof. An essentially similar argument can be applied to screw dislocations. A little thought will show that the passage of a twinning dislocation completely across a composition plane results not only in the conversion of a further layer of crystal of thickness x from matrix to twin, but also in the bodily relative movement of twin and matrix by an amount $b=(x$. $s)$. Thus it follows that a shear stress applied externally in the appropriate direction will cause twinning dislocations to move, i.e. will cause twins to grow.

In the particular case of a hexagonal lattice, with which we are primarily here concerned, one can readily calculate the magnitude of the twinning

dislocation vector. We have already given the expression for the twinning shear $s=\{(c/a)^2-3\}/\sqrt{3}\,.\,c/a$. The quantity x above is simply the spacing of $(10\bar{1}2)$ planes, and is thus equal to $ac\sqrt{3}/2\{c^2+3a^2\}^{1/2}$. The product of these two quantities gives the Burgers vector of the twinning dislocation as

$$(c^2-3a^2)/2(c^2+3a^2)^{1/2}$$

or

$$a\,.\,\{(c/a)^2-3\}/2\{(c/a)^2+3\}^{1/2}.$$

For Cd ($c/a=1{\cdot}885$) this has the value of $0{\cdot}108\,a$; for Zn ($c/a=1{\cdot}856$) it is $0{\cdot}088\,a$; while for Mg ($c/a=1{\cdot}624$) it is $-0{\cdot}076\,a$.

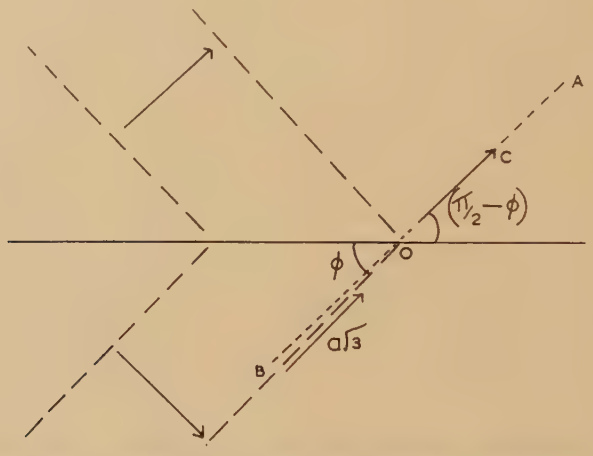
Before proceeding further, it will be convenient to discuss briefly the types of ordinary dislocations which can exist in hexagonal lattices. For dislocation loops lying in the (close packed) basal plane, and with their Burgers vectors also in this plane the situation is closely similar to that existing in face centred cubic lattices. Three equivalent elementary Burgers vectors are possible, of length a and lying in the directions $[1\bar{2}10]$ $[\bar{2}110]$ $[11\bar{2}0]$. Each of these can, and probably does, dissociate into two half dislocations (see Heidenreich and Shockley 1948). These are the dislocations responsible for the familiar basal plane glide in hexagonal crystals. It is not however topologically necessary that such dislocation loops should be confined to (0001) planes; they could be partially or wholly in prismatic or pyramidal planes of the type $\{1\bar{1}0\ n\}$ for example. But if one makes a model of the structure using hard spheres it soon becomes obvious that even if such dislocation loops do exist it would be difficult for them to move. The requisite movement of a typical sphere from its initial to its final equilibrium position, is manifestly through intermediate configurations of such high energy, that very large shear stresses would be necessary to bring it about, and a crystal would probably deform by some other mechanism before such stresses could be attained. A possible exception is the $(1\bar{1}01)$ plane: in this case an intermediate stable configuration exists which, although of greater energy than the initial and final states, does suggest the possibility of the existence of half dislocations, and also that the critical stress for slip on these planes may not be prohibitively high. It is interesting to note that this is the only slip system, other than basal plane slip, which is on record as having been observed in hexagonal metals. (Magnesium, see Bakarian and Mathewson (1943).)

We must, however, also consider the possibility of dislocations existing with a Burgers vector equal to c. For convenience these will be referred to as ' major ' dislocations to distinguish them from those whose Burgers vector ($=a$) lies in the basal plane (' minor ' dislocations). On the basis of recent theories of crystal growth (Frank 1949) one is led to postulate the existence of such major dislocations of screw character, to account for the growth habit of, for example, zinc, which condenses from the vapour in the form of flat plates, normal to the c-axis. But here again, by considering the behaviour of a hard sphere model, one is led to believe that although such dislocations may exist in the lattice, they will be unlikely to move under the action of the usual applied stresses. No

suggestion of any slip process which could be attributed to their presence is to be found in the literature.

One further point may be noticed : it is, as a general principle, possible for several dislocation lines to meet at a node, provided one condition is satisfied—namely that the vector sum of all the Burgers vectors as seen looking out from the node, is zero. This is a most valuable concept in face centred cubic lattices. But in hexagonal lattices the two types of Burgers vector (major and minor) are necessarily at right angles to one another. We can deduce at once that the two types of dislocation must form quite independent systems; and that although networks of minor dislocations are possible, with three-fold nodes, major dislocations can exist only as single loops.

Fig. 8



Let us now turn to consider the process of twin formation in a crystal which already contains dislocations. The interesting case is that of a major dislocation of screw character. It would be along a line such as OA (fig. 8) with its Burgers vector ($=c$) in the same direction. When the twin boundary has advanced to the position shown in the figure, what has become of the singularity in the crystal which previously lay along the line BO ? Can it have been converted into any of the permitted dislocations of the twin lattice ? The line BO will lie almost but not quite, in the basal plane of the twin. Consider two minor dislocations in the twin lattice both lying along BO, one with its Burgers vector dipping down into the plane of the paper at 30°, the other with its vector pointing upwards at 30°. Both of these are permitted directions in the twin lattice. The resultant vector of the double dislocation will be $a\sqrt{3}$ in the direction BO. This is almost the required answer, since $c \approx a\sqrt{3}$ for any real hexagonal lattice. But let us evaluate the difference between the two. The angle ϕ between the basal plane and the twinning plane is $\tan^{-1}(c/a\sqrt{3})$. Using this result it can readily be shown that the residue when the vector c in the matrix is transformed into $a\sqrt{3}$ in the twin is a vector lying in the twin boundary, and of magnitude

$a \cdot \{(c/a)^2-3\}/\{(c/a)^2+3\}^{1/2}$. This is exactly twice the value previously found for the Burgers vector of the twinning dislocation.*

We are thus led to the result that whenever a major dislocation existing in one member of a twin pair, crosses a twin boundary, it may be transformed into two minor dislocations in the lattice of the other member of the twin pair, together with two twinning dislocations. (With the aid of a stereographic projection showing the direction of the Burgers vectors in both members of the twin, one can see at once that no other simple relation of this kind is likely to exist.) This then provides a mechanism for the generation of twinning dislocations. But we may proceed one step further. It is well known that a dislocation line cannot end within the lattice, and we have already pointed out that there can be no interconnection between major and minor dislocations existing in the original lattice. Thus major dislocations must exist as closed loops (or else pass right through the crystal from one face to another). The twin boundary will cut such a loop of a major dislocation in two places (at least) and a moments consideration of the advance of the boundary from the instant when it first intersected the loop shows that these two points will be joined by two twinning dislocations constrained to lie in the twin plane. The situation is, in fact, the exact analogue of the Frank–Read ' source of dislocations ' and will behave in just the same way. The ends of the twinning dislocations are firmly anchored to the points where the major dislocation loop cuts the (present) position of the twin interface. But under the action of an applied shear stress of sufficient magnitude the rest of the line can form a loop of dislocation of increasing size, which will eventually spread over the whole of the twin interface, adding one more layer to the twinned region, and nevertheless still remaining in existence to repeat the whole process indefinitely.

On this view, then, a number of the familiar problems of dislocation theory, as applied to slip processes, will have their analogues in the growth of twins. The twinning problem is in some ways the simpler, since from its nature one and only one twinning dislocation can pass across every plane of the lattice (counted normal to the twinning plane). For example, the theory which accounts for a yield point in terms of an atmosphere of impurity atoms formed around dislocations might be applied also to twinning dislocations. It would predict a critical resolved shear stress law for the growth of a twin. The critical stress is inversely proportional to the Burgers vector (Cottrell and Bilby 1949) and so would be considerably greater for twinning than for slip. This is in qualitative and rough quantitative agreement with our results on cadmium, and those of Davidenkov *et al.* on zinc, if we assume that the critical twinning stress which is observed experimentally corresponds to the growth of existing twin nuclei. This is *à priori* not improbable (Oliver 1952).

It is not difficult to invent detailed mechanisms to account qualitatively for the other observations reported in § 3 but at present these must

* This idea is due in the first place to Dr. F. C. Frank to whom we are also grateful for many helpful discussions of these matters.

remain somewhat speculative. It is clear that more experiments are needed, and it may well be that they will throw new light on some of the older problems of plastic deformation by the movement of 'ordinary' dislocations.

[While this paper has been in preparation, Cottrell and Bilby (1951) have published an account of a general theory of the growth of deformation twins which puts forward suggestions essentially similar to the above. They consider in detail its application to body-centred and face-centred cubic lattices. It may be noted that the four conditions which they lay down (p. 576) are all satisfied by the mechanism we propose here. Three of them have already been discussed, and the fourth (No. 2) can readily be verified thus: The Burgers vector of the 'pole' dislocation in our model$=c$; the component of this normal to the composition plane$=c\cos\phi=ac\sqrt{3}/(c^2+3a^2)^{1/2}$. This is twice the expression already given for the repeat distance normal to the twinning plane—previously denoted by x. This satisfies Cottrell and Bilby's second condition if we remember that our twinning dislocations are necessarily produced in pairs.]

Acknowledgments

We would like to express our thanks to colleagues in the laboratory, particularly Dr. G. Wyllie and Dr. F. C. Frank for many helpful discussions. One of us (D. J. M.) also wishes to acknowledge financial assistance from the Department of Scientific and Industrial Research.

References

Andrade, E. N. da C., and Roscoe, R., 1937, *Proc. Phys. Soc.*, **49,** 152.
Bakarian, P. W., and Mathewson, C. H., 1943, *Trans. A.I.M.E.*, **152,** 226.
Barrett, C. S., 1943, *Structure of Metals* (New York: M'Graw Hill) p. 174; 1949, *Cold Working of Metals* (Cleveland: A.S.M.) pp. 78 *et seq.*
Boas, W., and Schmidt, E., 1929, *Z. Physik.*, **54,** 16.
Cottrell, A. H., and Bilby, B. A., 1949, *Proc. Phys. Soc.*, **62,** 49; 1951, *Phil. Mag.*, **42,** 573.
Cottrell, A. H., and Gibbons, D. F., 1948, *Nature, Lond.*, **162,** 488.
Davidenkov, N. N., Kolesnikov, A. F., and Federov, K. N., 1933, *J. Exp. Theor. Phys. USSR*, **3,** 350.
Frank, F. C., 1949, *Conference on Crystal Growth.* (London: The Faraday Society) p. 48; 1951, *Phil .Mag.*, **42,** 809.
Frank, F. C., and van der Merwe, H. J., 1949, *Proc. Roy. Soc.* A, **198,** 205.
Gough, H. J., and Cox, H. L., 1929, *Proc. Roy. Soc.* A, **123,** 143; 1930, *Ibid.*, **127,** 453.
Heidenreich, R. D., and Shockley, W., 1940, *Conference on the Strength of Solids.* (London: The Physical Society) p. 57.
Kolesnikov, G., 1933, *Phys. Zeit. Sow. Union.*, **4,** 651.
Mathewson, C. H., 1928, *Trans. A.I.M.E.*, **78,** 7.
Mathewson, C. H., and Phillips, A. J., 1927, *Trans. A.I.M.E.*, **74,** 143; 1928, *Ibid.*, **78,** 445.
Millard, D. J., 1950, *Ph.D. Thesis.* Bristol University.
Miller, R. F., 1936, *Trans. A.I.M.E.*, **122,** 176.
Oliver, D. S., 1952, *Research*, **5,** 45.
Schmidt, E., and Boas, W., 1935, *Kristallplastizität* (Berlin: Springer) p. 100, English Translation (Hughes: London 1950) p. 94.

XXXIX. *Further Observations of Growth Patterns on Silicon Carbide (Si–C) Crystals*

By Ajit Ram Verma
Royal Holloway College, University of London*

[Received November 2, 1951, amended December 31, 1951]

Abstract

This paper reports further observations of growth patterns on silicon carbide crystal faces using techniques previously described. In addition to the simple types of spirals in which the step height was shown to be equal to the height of an x-ray unit cell, spirals originating from dislocations of multiple strength, the step height being a multiple of the height of an x-ray unit cell, are reported. More complex geometrical growth patterns of groups of dislocations are illustrated. The conditions for the interaction of growth fronts issuing from a number of dislocations are discussed and shapes of the loci of points of intersection of such growth fronts are calculated and illustrated. It is observed that in any one region of the crystal, the screw dislocations are not only predominantly of the same sign but have also the same strength.

§1. Introduction

Observations of ' growth spirals ' on silicon carbide (Si–C) crystals have been reported earlier (Verma 1951) affording experimental support for the theory of growth of crystals depending upon the presence of dislocations (Burton, Cabrera and Frank 1951). The present paper deals with further observations on the more complex geometrical growth patterns of groups of dislocations and the interaction of a number of dislocations. Also spirals originating from dislocations of multiple strength are reported.

§2. Spirals Originating from Dislocations of Multiple Strength

The step formed by the emergence of a screw dislocation on to the crystal face will have a height corresponding to the differences in the amounts of slip between neighbouring regions of the crystal. Because of the existence of discrete atomic structure in the glide plane, this will be limited to the value of the x-ray unit cell of the crystal or in some lattices a small multiple or fraction of this. The spirals reported earlier (Verma 1951) were mostly of the simple type and the measured step height was shown to be equal to the height of the x-ray unit cell (in type II = 15 Å). Because of the small step height visibility was low.

Some of the spirals now observed have a much higher visibility and interferometric measurement shows that they have a relatively large step height ; indeed, they can often be seen simply with a microscope using

* Communicated by Professor S. Tolansky.

bright field illumination without the need for silvering the crystal. The figures reproduced here have been taken by the technique used previously (Verma 1951).

As silicon carbide (Si–C) is a polytypic crystal, a large step height may be due to its being equal to (1) the height of a large x-ray unit cell, (2) a multiple of a smaller x-ray unit cell. The latter type of spiral results from a dislocation of multiple strength, the Burgers' vector being a multiple of the x-ray unit cell.

Fig. 1 (Plate XVIII*) illustrates such a spiral originating from a dislocation of multiple strength, observed on a crystal of type II (determined goniometrically). The step height between the successive arms of the spiral is nearly 620 Å as determined by the shift of the multiple-beam Fizeau fringes over these steps. The rate of advance of a multiple step will be controlled by the deposition rate at the bottom of the step (Frank 1951) and as long as the bottom of the step is not privileged with respect to the diffusion of the molecules from the gas, the multiple step will not dissociate into its component steps. In fig. 1 (Plate XVIII) the edges do not remain straight after three or four turns from the centre and become irregular. This behaviour continues for a few more turns after which the steps dissociate into five visible components. The edges of these multiple steps will not be close packed. The Fizeau fringes are observed to be continuous over the edges of the steps while climbing them. This indicates that the edges are not steep and vertical. Furthermore the crystal face between the successive edges is not close packed. This is shown in fig. 2 (Plate XVIII) where some faint step lines can be seen between successive edges.

The dissociation of a step into its components is seen to take place preferentially in certain orientations. In figs. 1 and 2 (Plate XVIII) it is seen that the dissociation has taken place preferentially in one orientation and at 60° to this direction this dissociation is much less marked. Fig. 3 (Plate XVIII) is another example showing this better, where the dissociation is observed on the three alternate sides of the hexagonal spiral. This is easily comprehensible since in Si–C a growth layer which is fastest in any one orientation becomes slowest in orientations at 60° to it ; the faster moving steps will overtake it and pile up behind it.

Dissociation is seen to take place quite near the origin for a further example (fig. 4, Plate XVIII). Here amongst the dissociated steps, in three orientations the outermost edges (which will correspond to the bottom layers of the group) are straight and parallel to each other in the successive arms of the spiral, whereas in the inner edges the usual irregularity is observed.

Fig. 5 (Plate XVIII) shows a case of multiple dislocations. In contrast to fig. 4 (Plate XVIII) the growth is faster at the six corners, the rate increasing towards the outer (or bottom) layer in the group. At some of these corners crosslacing of the step lines is observed and this might be responsible for the faster growth, or alternatively there may be other explanations.

* For plates see end of issue.

§3. Pairs of Dislocations of Opposite Sign

When two dislocations of opposite sign emerge on the crystal face, they attract each other with a force which is inversely proportional to the distance between them and increases with the strength of the dislocation (Taylor 1934). There is however a constraining force anchoring these dislocations to their equilibrium positions between the lattice rows, and this force depends upon the elastic constants of the medium (Peierls 1940, Nabarro 1947). This force should also depend in some way upon the strength of the dislocation. At a certain distance of separation between the two dislocations these forces will balance each other. During the process of growth of Si–C, when the temperature is high, the two dislocations of opposite sign will draw towards one another and will finally coalesce, joining to form a common hollow core, unless the distance of separation is more than the above equilibrium value. Experimentally, pairs of dislocations of opposite sign are seldom observed. Amongst all such pairs observed so far on Si–C, a pair of dislocations of opposite sign close to each other such that the two spirals are equally developed (i.e. the first, second etc. arms of the spiral from one dislocation meet the corresponding arms from the other) has not yet been observed for the dislocations of elementary Burgers vector (e.g. 15 Å for type II). However for a comparatively large Burgers vector one pair of closely spaced and equally developed dislocations has been observed. Fig. 7 (Plate XVIII) shows this rare case in which the step height is nearly 120 Å as measured by the shift of the multiple-beam Fizeau fringes over these steps (fig. 6, Plate XVIII).

Fig. 8 (Plate XIX), which is a bright-field photograph of an unsilvered crystal, shows two dislocations of opposite sign each giving rise to a hexagonal spiral. When the successive arms of the two spirals meet each other they generate closed loops. At the points of intersection an edge can be observed, indicating that the levels of the two steps joining together are not the same.

Another rare case is shown in fig. 9 (Plate XIX) which consists of closed hexagonal sheets with no trace of the dislocations at the centre. The closed hexagonal figures might have originated from two dislocations of opposite sign, alternatively there may be some other mechanism. (No new closed loops are generated at the centre after the dislocations are drawn close to each other compared to ρ_c.)

The possible 'neutralization' of pairs of dislocations of opposite sign will leave only the surplus of one kind to be observed. This may explain the observation that in any one region of the crystal the dislocations are predominantly of one type. Further, after this 'neutralization' the surplus left behind will be much less than the original value. (For N dislocations randomly formed, the probable excess will be $\sim N^{\frac{1}{2}}$.) Thus the observed maximum density of dislocations $\sim 10^5/\text{cm}^2$ (in fig. 18 (Plate XX), all of the same type) is much less than the theoretically predicted value $\sim 10^8/\text{cm}^2$.

§4. Groups of Dislocations

Quite often groups of dislocations arranged in different ways emerge to the surface of the crystal. A group of some 50 dislocations of the same sign arranged along a line of length L is seen in fig. 10 (Plate XIX), to generate a spiral system of s branches. The resultant activity of the group will be $s/(1+L/2\pi\rho_c)$ times that of a single dislocation (Burton, Cabrera and Frank 1951) where ρ_c is the radius of the critical nucleus. Fig. 11 (Plate XIX) illustrates another such group of dislocations where steps are seen to group together giving a repeat pattern.

In the general case of a group of like dislocations Burton, Cabrera and Frank (1951) predict that a pit may be developed on the crystal face if the growth fronts have difficulty in penetrating the group itself. This prediction is confirmed by the case illustrated in fig. 12 (Plate XIX) where over 50 circular step-lines originate from an elliptical hole. Obviously at such a source pairs of dislocations of opposite sign will neutralize each other leaving the excess of only one type to promote growth.

§5. Growth from a Number of Interacting Dislocations

Interaction of Growth Fronts and Curves of Intersection

Very often a group of dislocations, instead of originating from one source, spread on the surface of the crystal. The growth fronts spreading from the various sources meet each other producing complicated growth patterns in acccordance with their properties. The general principles that govern their interaction and the shape of the locus of points of intersection are as follows :

The successive arms of the spirals originating from any two dislocations can go on joining with each other only if the two dislocations are of the same strength, i.e. the two step heights are equal. Conversely, if successive steps from two dislocations join and fuse together, it can be concluded that the step heights of the two spirals are equal. Since this joining together of steps is observed in figs. 15 and 18 (Plate XX) for the successive steps from all the interacting dislocations, the step height must be a fixed unit and its measurement on any one of them will give the value of all the rest. Thus in any one region of the crystal, the dislocations are not only predominantly of the same sign but also have the same strength.

If then the nth step from one source O_1 meets the mth step from the other source O_2, they will join together if they are at the same level. Thereafter pairs of steps that can join together are $(n+1)$th with $(m+1)$th ; $(n+2)$th with $(m+2)$th etc. such that the order difference between these steps fusing with one another is always constant and equal to $(n-m)$. Macroscopically speaking, on either side of the locus of points of intersection, the portions of the two pyramids should slope in the same direction, i.e. on travelling along this locus of points of intersection we will be moving from a higher to a lower level, or vice versa, for both the pyramids. Examples of this can be seen in fig. 15 (Plate XX).

Further, if the spacing between the nth and the $(n+1)$th arms of one spiral is d_1 and between the mth and the $(m+1)$th arms of the other is d_2, the two pairs of steps will meet the locus of points of intersection at angles θ_1 and θ_2 respectively, such that $\sin\theta_1/\sin\theta_2 = d_1/d_2$. The value of $(\theta_1+\theta_2)$ at a particular point of intersection is fixed by the initial conditions of the distance of this point from the two sources, the distance between the two sources, and the shapes of the two spirals.

In the more general case d_1 and d_2 are the lengths of perpendiculars dropped from the point where the $(n+1)$th and $(m+1)$th arms meet each other, upon the tangents drawn to the nth and mth arms of the two spirals at the point where they meet. The tangents to the nth and $(n+1)$th arms of one spiral may not be parallel to each other, so that not only the spacing between the successive arms changes, but also their directions may change, in which case the locus of points of intersection will have to turn suitably in order to satisfy the above sine condition. This is illustrated in fig. 13 (Plate XIX) where with the change of direction by the steps through an angle about 60°, the locus of points of intersection also turns through a similar angle. Other examples of the interaction of growth fronts and the loci of their points of intersection in accordance with the above conditions are seen in figs. 15 and 18 (Plate XX), where a large number of polygonal spirals with varying spacings are interacting with one another. Another example of the interaction is seen in fig. 14 (Plate XIX), where as a result of the meeting of the growth fronts from three dislocations closed triangular loops are formed. (The figure includes only two dislocation centres. The third one will be on the left side.)

If the two series of steps originating from two dislocations, on reaching the same region of the crystal, are unable to satisfy any of the above conditions, e.g. sine condition, they will not join with one another. These steps will then terminate in such a region. The points where the steps terminate will be at the ends of screw dislocations or at discontinuities in level. Such behaviour is seen in fig. 15 (Plate XX) along a line running from left to right nearly in the middle of the figure, with a small gap in the centre, over which region four steps of the two spirals are seen to join together. We can reach any one of these steps (say the second from the left) from a point A in a number of ways. In going from A to the chosen step we have to go down 30 steps, this number being independent of the path taken. A direct path can be traversed or one involving the crossing of different hexagonal spirals by going round the right hand end of the line of discontinuity. This is what is expected since the different spirals are of the same strength.

§6. Special Case : Circular Spirals

By introducing the approximation used for the interferometric measurements of step heights, according to which these spirals may be regarded as composed of circles of constant spacing $\delta r = 4\pi\rho_c$, the condition $(n-m)=$ constant directly gives the locus of the point of intersection P of the two

rounded spirals originating from the two source centres O_1 and O_2. Thus $(n-m)\,.\,\delta r$ will be a constant from which we have $O_1P-O_2P=$constant. This gives the locus of the point of intersection, a hyperbola which reduces to a straight line perpendicular to O_1O_2 and passes through its middle point when $n-m=0$ (Burton, Cabrera and Frank 1951).

A locus of points of intersection, which is hyperbolic, is illustrated in fig. 16 (Plate XX) for two circular spirals unequally developed. The condition $(n-m)=0$ corresponds to the symmetrical case of two spirals equally developed and is illustrated in fig. 17 (Plate XX) where the locus of points of intersection is seen to be approximately a straight line, perpendicular to the line joining the two dislocations and passing through the middle point. Here the two series of steps with equal spacing (i.e. $d_1=d_2$) meet the locus of points of intersection at equal angles ($\theta_1=\theta_2\neq 0$). In fig. 16 (Plate XX), as the spacings of the two spirals become equal (i.e. $d_1=d_2$), in accordance with the sine condition $\theta_1=\theta_2$, and in this case the steps tend to be parallel to each other at the points of intersection (i.e. $\theta_1=\theta_2=0$).

As shown by Burton, Cabrera and Frank (1951) a small influence is transmitted along each step from the points where the two spirals meet, into the respective centres. This will tend to increase the rate of rotation of the spiral whose centre is nearer the points of contact (the upper one in fig. 16, Plate XX), trying to synchronize the two in phase. This will smooth out the corners formed by the meeting of the two series of steps. This is observed in all the interacting spirals.

Acknowledgments

I am grateful to Professor S. Tolansky for his interest and encouragement in the course of the work. Thanks are due to Mr. H. Rahbek for useful discussion. I am indebted to Dr. E. W. J. Mitchell for a further supply of silicon carbide crystals. This work has been carried out during the tenure of a scholarship from the British Council while on study leave from Delhi University.

References

Burton, Cabrera and Frank, 1951, *Phil. Trans. Roy. Soc.*, **243**, 299.
Frank, F. C., 1951, *Phil. Mag.*, **42**, 1014.
Nabarro, F. R. N., 1947, *Proc. Phys. Soc.*, **59**, 256.
Peierls, R., 1940, *Proc. Phys. Soc.*, **52**, 34.
Taylor, G. I., 1934, *Proc. Roy. Soc.* A, **145**, 362.
Verma, A. R., 1951, *Phil. Mag.*, **42**, 1005.

XL. *The Origin of Cosmic Rays in Solar or Stellar Disturbances*

By L. Riddiford
Physics Department
and
S. T. Butler*
Mathematical Physics Department, University of Birmingham†

[Received February 5, 1952]

Abstract

Charged particles may attain cosmic-ray energy by betatron acceleration in trochoid-like orbits in the varying magnetic fields which occur in star spots. For sunspots, cyclic acceleration produces maximum energies of about 10^{13} ev for protons and 10^{15} ev for heavier particles. If the source of ions is constant, the differential number-energy spectrum of the relativistic particles has the form of an inverse power law. The power in this law is the same for protons and heavier particles, and lies between 1·7 and 3·0 independently of assumptions about the constants of the spot.

§1. Introduction

It is many years since Swann (1933) made the proposal that particles of energy up to 10^9 ev may be produced by acceleration in the electric fields associated with the changing magnetic fields in sunspots. Swann did not consider in detail the possibility of cyclic orbits, and he concluded that the cosmic ray particles striking the earth, which must possess energies in excess of 10^9 ev to traverse the earth's magnetic field without substantial deflection, must, if accelerated by this mechanism, come from stars other than the sun.

We shall show that, in a magnetic field constant in space but rising linearly with time, ions will generally circulate in trochoid-like orbits with a net gain of energy in each cycle. With reasonable values for the sunspot constants, proton energies in excess of 10^{12} ev are obtainable. Further, particles which start at different times and at different places in the sunspot will attain different energies by the time the sunspot field has reached its maximum value. With a constant source of particles, the acceleration leads quite naturally to an inverse power law for the differential number-energy spectrum, the value of the power being in good agreement with the experimental one. Since phenomena like sunspots presumably occur in other stars, our results do not necessarily point to a solar origin of the cosmic radiation. It is evident from a recent review of the origin of cosmic rays by Terletski (1949) that the exact origin is not clearly known.

* Now at the Laboratory for Nuclear Studies, Cornell University, Ithaca, New York.

† Communicated by Professor P. B. Moon, F.R.S.

§2. The Equations of Motion

There is little experimental evidence available on the shape and extent and time-rate of rise of the magnetic fields associated with sunspots, although it is known that the maximum value of the field may be as high as 4 000 gauss. However, the active regions with which spots are associated commonly have a radius of $2-3\times10^9$ cm, and in 1858 a very large spot of radius 2×10^{10} cm was observed. In many cases the disturbed region consists of a cluster of spots each of radius less than 10^9 cm. The magnetic field is almost certainly due to a current eddy, and since the field of a circular loop of current is roughly uniform throughout a cylinder of height and radius equal to the radius of the loop, it has been assumed that the field is uniform within the region of interest. It has also been assumed that the field increases linearly with time, measured from the instant when the field is zero. The results obtained do not depend strongly on either of these assumptions.

The polar equations of motion for a particle of charge e and mass $m=m_0\,(1-\beta^2)^{-1/2}$ moving at the position $(r,\,\phi)$ in a magnetic field $H=a_0t$ constant in space but increasing linearly with time are:

$$mr^2\dot{\phi}-\tfrac{1}{2}a_0er^2t=K\,, \qquad \ldots\ldots\ldots \quad (1)$$

$$\frac{\partial}{\partial t}(m\dot{r})+\tfrac{1}{4}\frac{a_0^2e^2rt^2}{m}=K^2/mr^3\,, \qquad \ldots\ldots\ldots \quad (2)$$

with K the constant of integration. The origin is at the centre of the sunspot.

§3. The Energy Obtainable by Cyclic Acceleration

For the particular case in which the particle starts from rest at time $t=0$ (which is the only case considered by Swann), $K=0$ and eqn. (2) has the solution

$$r=\left|\frac{A\sqrt{2}}{(a_0et)^{1/2}}\left\{\int^t\omega dt\right\}^{1/2}\left\{J_{1/4}\Big(\int\omega dt\Big)+\beta J_{-1/4}\Big(\int\omega dt\Big)\right\}\right|, \qquad \ldots \quad (3)$$

where $\omega=(a_0et)/(2m)$, and A, β are constants. $J_{1/4}(\int\omega dt)$ and $J_{-1/4}(\int\omega dt)$ are Bessel functions of the first kind in the argument $\int\omega dt$. The solution (3) holds exactly if $\dot{m}=0$, which is true for injection at non-relativistic velocity. For times later than $t=0$, $\dot{m}\neq0$, but at such times that the energy is sufficiently high to be of interest, $\omega t\gg1$ and the solution (3) is then also exact.

Consequently, by use of eqn. (3) an expression for the final energy at time t can be derived. This is

$$T=\frac{{m_0}^{1/4}c}{\pi^{1/2}}\cdot(a_0e)^{3/4}\cdot\Gamma(\tfrac{3}{4})\cdot r_0t^{1/2}\ \text{erg}, \qquad \ldots \quad (4)$$

where r_0 is the radius at the injection time $t_0=0$, and Γ is the gamma-function.

For a magnetic field rising to a maximum of 3 000 gauss in $t=10^4$ sec, $a_0=10^{-11}$ e.s.u. per second. Also $r_0=2\times10^9$ cm, so that according to eqn. (4) protons will attain an energy of $1{\cdot}7\times10^{12}$ ev. For a completely stripped barium atom, T is $9{\cdot}5\times10^{13}$ ev, and for a singly charged barium atom it is $5{\cdot}7\times10^{12}$ ev. For $r_0=2\times10^{10}$ cm, $T=1{\cdot}7\times10^{13}$ ev for protons and 10^{15} ev for a barium nucleus. Since T is proportional to $a_0^{3/4}t^{1/2}$, it does not depend strongly on a_0 or t. If the field rises to 3 000 gauss in $10^6(10^2)$ sec, then $a_0=10^{-13}(10^{-9})$ and the peak proton energy for $r_0=2\times10^9$ is $5{\cdot}2\times10^{11}$ $(5{\cdot}2\times10^{12})$ ev.

It seems possible then, that particles of cosmic ray energy are obtainable by betatron acceleration in sunspots. An investigation is now made of the solution for injection times t_0 other than zero, and injection energy not zero.

§4. Energy Dependence Upon Injection Time

Ions which enter or are formed within the sunspot at times when the magnetic field differs from zero are curled up into relatively small circles by the field. Their orbits will enclose less magnetic flux than the orbits of those entering at earlier times, they will be accelerated more slowly and not reach as high a final energy.

From eqn. (2), the radial acceleration of the particle is zero when

$$r_i^2=\frac{2\,|K|}{a_0 et}. \qquad \ldots\ldots\ldots \quad (5)$$

Thus r_i may be defined as the radius of the 'equilibrium orbit', and the 'instantaneous orbit' is one which oscillates about this.

Writing then

$$r=r_i+x, \qquad \ldots\ldots\ldots \quad (6)$$

and substituting into eqn. (2), to the first order in x/r_i

$$\frac{\partial}{\partial t}(m\dot{x})+\frac{a_0^2e^2xt^2}{m}=-\frac{\partial}{\partial t}(m\dot{r}_i). \qquad \ldots\ldots\ldots \quad (7)$$

The solution of eqn. (7) is the sum of a complementary function and a particular integral. The former can be written down by analogy with eqn. (3), and the latter may be either a power series in ωt or an asymptotic series valid for $\omega t\gg1$. For the region $\omega_0t_0>10$ the solution has the asymptotic form

$$x=\frac{B}{(a_0et)^{1/2}}\cdot\sin\left\{\int_0^t\omega dt+\gamma\right\}-\tfrac{3}{4}\frac{r_i}{\omega^2t^2}, \qquad \ldots\ldots \quad (8)$$

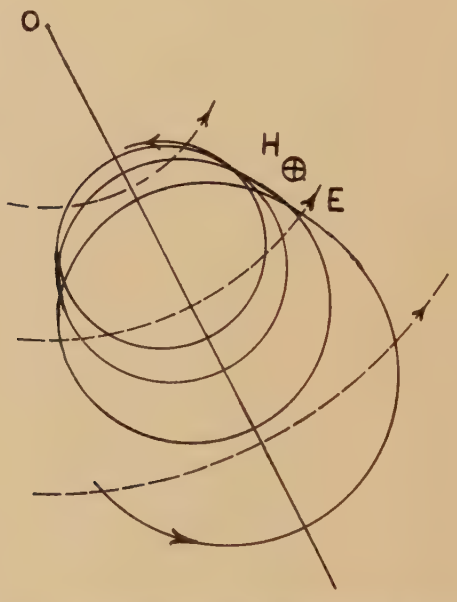
where B, γ are constants and now $\omega=(a_0et)/(m)$. ω_0 is the value of ω at the injection time t_0. This solution permits development of the theory in an analytic form, whereas the more exact equation for x, valid for $\omega_0t_0<10$, necessitates numerical calculations involving series containing as many as fifteen terms.

Since x is an oscillatory function, and for late injection times the path of the particle is curled up in the crossed electric and magnetic fields, the orbit is trochoidal. This may be shown in another way. If $u_0 = r_0\dot{\phi}_0$ is the tangential injection velocity, from eqn. (1) K is negative provided $u_0/\omega_0 r_0 < 1$. Then by use of eqn. (5), eqn. (1) can be written in the form

$$mr^2\dot{\phi} \simeq a_0 e t r_i x. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (9)$$

The inequality is equivalent to $x/r_i < 1$, and it may be shown that $x/r_i \ll 1$ for all times of interest. $\dot{\phi}$ then is also an oscillatory function, and the orbit of the particle never loops the origin. This is important for it reduces the chance that the accelerating particles will form a secondary current

Fig. 1



looping the origin of the sunspot and producing an induced magnetic field of sufficient strength to annul the primary field (Alfven 1950). The current of particles on the outer half of each loop of a possible orbit will be cancelled by the current on the inner half of the loop.

That the equilibrium orbit moves (in the plane of the paper) as drawn schematically in fig. 1, may be shown analytically once the final result that $m \propto t^{1/2}$ (eqns. (13) and (14)) is known. In fig. 1 the magnetic field is normal to the paper, and the lines of electric force are circles about the sunspot centre O. The amplitude of the oscillatory motion is damped as $t^{-1/2}$, or for the linearly rising field considered here, as $H^{-1/2}$, which is the manner in which the free oscillations of a conventional induction accelerator, first described by Kerst and Serber (1941), are damped. Hence the amplitude of the trochoid decreases with time. Its centre moves in a

straight line towards the centre of the spot along the radius vector on which the particle originated.

It is apparent that the particles will alternately gain and lose energy depending on whether $\dot{\phi}$ is positive or negative. But the amount gained on one half-cycle will always exceed the amount lost on the ensuing half-cycle because the particles are being continually curled up in the increasing magnetic field so that their orbits enclose less magnetic flux. This results in a steady increase in energy according to the first order calculations presented here.

The velocity v of the particle is given to a very close approximation by the equation

$$v^2=\dot{x}^2+\omega^2x^2 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (10)$$

For $\omega_0 t_0>10$, using eqn. (8) and the equations $x_0=u_0/\omega_0$ and $\dot{x}_0=u_0+r_0/2t_0$, this becomes

$$v^2=\frac{\omega^2}{\omega_0{}^2}\cdot\frac{t_0}{t}\left[\left\{\left(u_0+\frac{r_0}{2t_0}\right)-\frac{3}{4}\cdot\frac{r_0}{t_0}\cdot\frac{1}{\omega_0{}^2t_0{}^2}\right\}^2+\left\{u_0+\frac{3}{4}\cdot\frac{r_0}{\omega_0{}^2t_0{}^2}\right\}^2\right]. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (11)$$

To good approximation eqn. (11) may be written

$$v=\frac{\omega}{\omega_0}\left(\frac{t_0}{t}\right)^{1/2}\left\{\left(u_0+\frac{r_0}{2t_0}\right)^2+u_0{}^2\right\}^{1/2} \quad . \quad . \quad . \quad . \quad . \quad . \quad (12)$$

From this equation it follows that the energy $T=mvc$ of an extreme relativistic particle is, for $u_0\gg r_0/2t_0$,

$$T=\sqrt{(2)}m_0u_0c\,(t/t_0)^{1/2}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (13)$$

and for $u_0\ll r_0/2t_0$,

$$T=\frac{m_0r_0c}{2}\cdot\frac{t^{1/2}}{t_0{}^{3/2}}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (14)$$

For $u_0=0$, $r_0=2\times10^9$, $t=10^4$, $H=3\,000$, and $\omega_0t_0=10$, $T=2{\cdot}28\times10^{11}$ ev for protons. Previously it was shown that for $\omega_0t_0=0$, $T=1{\cdot}7\times10^{12}$ ev. We have also carried out numerical calculations for the range $0<\omega_0t_0<10$, and the results are as follows:

ω_0t_0	=0·002	0·01	0·1	1·0	2·0	3·0	4·0	5·0	7·0	9·0	11·0	13·0
T (ev)=	2·02	1·88	1·38	7·52	5·59	4·53	3·85	3·39	2·75	2·31	2·01	1·82
	$\times10^{12}$	$\times10^{12}$	$\times10^{12}$	$\times10^{11}$	$\times10^{11}$	$\times10^{11}$	$\times10^{11}$	$\times10^{11}$	$\times10^{11}$	$\times10^{11}$	$\times10^{11}$	$\times10^{11}$

The agreement between the values obtained analytically and numerically demonstrates the validity of the calculations. Eqns. (13) and (14) hold for any particle injected at non-relativistic velocity, the final energy being proportional to its mass. Only those particles which are injected during the relatively short interval of time $0<t<$ say 10 secs, are accelerated to cosmic ray energies.

It is important to notice that according to eqns. (13) and (14) T is independent of a_0. In eqn. (13) T is independent of r_0, and in eqn. (14) it is proportional to r_0. Consequently, the result is independent of the time-rate of rise of field, and does not depend strongly on the place of origin of the particle in the sunspot.

§5. The Energy Spectrum

Since energies of experimental interest (10^9—10^{11} ev) are attained by particles starting at times such that $\omega_0 t_0 > 10$, a theoretical energy spectrum may be obtained from eqns. (13) and (14). Let $f(T)dT$ be the number of particles with energy between T and $T+dT$. If there exists a source of ions constant during the relatively short injection time, then

$$f(T) \propto 1 \Big/ \frac{dT}{dt_0}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (15)$$

Consequently, from eqns. (13) and (14), for $u_0 \gg r_0/2t_0$

$$f(T) \propto \frac{1}{T^3}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (16)$$

and for $u_0 \ll r_0/2t_0$

$$f(T) \propto \frac{1}{T^{5/3}}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (17)$$

To obtain eqn. (17) we have ignored the dependence of T on r_0 (eqn. (14)). This is justifiable since it is probable that most of the particles originate near the edge of the spot. Thus for extreme relativistic particles the power of T cannot lie outside the range 1·7 and 3·0, independent of any sunspot parameter and also of the mass of the particle. The experimental value for primary rays is known to be about 2·5 for particles of any mass (Bradt and Peters (1950)), values ranging from 1·8 to 2·7 having been reported (Hilberry (1941), Winckler, Stix, Dwight and Sabin (1950), van Allen and Singer (1950)). The experimental differential number-energy spectrum of the cosmic rays at sea-level has a maximum at the low energy end in the region of 10^9 ev. There exists a threshold at about this energy due to the earth's magnetic field (Lemaitre and Vallarta (1936)), and the sharpness of the maximum in the theoretical curve resulting from a correction for this depends upon what range of directions is being considered for the direction of the incoming rays. Further, it is essential for the experimenter to record only incoming rays and not secondary particles from surrounding objects. Consequently, care is necessary in comparing theoretical and experimental curves at the low energy end of the spectrum. In fact, however, as pointed out by Neher (Wilson (1951 a)), the high altitude measurements indicate no such maximum, although recently there has been evidence for a decrease in the slope of the spectral curve of the primary rays at low energies just above the geomagnetic threshold (Winckler *et al.* (1950), Van Allen (1950)). According to the present theory, for the non-relativistic case ($T < 10^9$ ev) the index in the power law must lie within the range 1·3—2·0, being lower than in the relativistic region ($T > 10^9$ ev). The present theory predicts no maximum, although it is quite possible that a suitable refinement of it may well indicate one. For instance, those particles which are injected at later times are accelerated more slowly and are less able to overcome the ionization and other losses and so be fully accelerated.

The power law holds to the highest energy attainable, which has been shown to be 10^{13} ev for protons and 10^{15} ev for heavier particles accelerated in sunspots. In any event the Richtmeyer–Teller trapping magnetic field (Richtmeyer and Teller (1949), Alfven (1949)) will not retain protons of energy higher than 10^{13} ev cycling in confined orbits within the solar system, so that those more energetic protons which may enter the solar system from other stars would have to strike the earth on their only passage through that system to be observed. This suggests that the extensive showers are due to the heavier particles, unless very energetic protons are produced in spots on other stars. The spectrum may be distorted somewhat by phenomena such as nuclear collisions which occur either in the sunspot or in interstellar space. These would tend to increase the slope of the spectral curve.

§6. Discussion

The gas in the photosphere and chromosphere, being mainly hydrogen at 5 000–6 000° C, will not be completely ionized (Saha (1920)). Consequently, charged particles will not be accelerated to very high energies unless the induced electric field is greater than the energy lost per centimetre by ionization and collisions with free electrons, and other causes. For protons the ionization loss curve has a maximum at about 100 kev. According to calculation (Blanch, Lowan, Marshak and Bethe (1941)), the gas pressure falls off very rapidly with height in the chromosphere, and so there will exist a critical level above which ions of any energy will be accelerated, and below which only ions starting with an energy greater than 100 kev will be accelerated. This level will not be sharply defined, for the gas will be quickly ionized by secondary collision as soon as the accelerating electric field is set up. For an electric field of order 1 volt per cm, as is involved here, the critical radius, where the gas pressure is about 10^{-3} mm of mercury, is situated deep in the chromosphere, just above the photosphere where the sunspots occur.

For acceleration to take place in the high pressure region it is necessary to postulate a source of high energy ions, which could reasonably be assumed to be a constant one. These may be produced by local electric discharges in a relatively intense field, or thermonuclear reactions. The secondary ions produced in this region will not be fully accelerated.

All ions initially present in the low pressure region will be accelerated. In this region ionization will be complete in a time short compared to the injection time (during which particles are accepted for acceleration to energies in excess of 10^9 ev). The secondary ions are all formed shortly after $t=0$ and will have energies approaching the maximum obtainable from the spot. The amount of gas which will stream into the spot during the acceptance time can hardly exceed that initially present, so that the prospect of obtaining a satisfactory energy spectrum is small. Blackett (1950) has raised a serious objection to the acceleration of all ions on the grounds of an elementary energy consideration, which applies, however, to a closed box in which the energy acquired during acceleration is derived from the thermal kinetic energy of the particles in the box. In the present

mechanism the dimensions of the generator providing the energy may exceed by several orders of magnitude the dimensions of the box of particles being accelerated, and further, the density of particles in the generator may be several orders of magnitude greater than in the box.

The plane in which the ions are being accelerated will tend to move away from the sunspot in a direction parallel to its axis, to regions of weaker magnetic field, during the acceleration time. It is essential that the shape of the field be such that the particles do not leave the spot before acceleration to cosmic ray energy is complete, and also that they do not remain sufficiently long to be decelerated as the field returns to zero. Since the particles receive most of their energy in the early stages of acceleration ($m \propto t^{1/2}$), and further the rise-time of the magnetic field may well be short compared to its duration (Hoyle (1949 a)), there will be plenty of time for extraction, and it may be possible to meet these objections to the theory. To support the arguments already given, for an injection mechanism, we may suppose that the ions accelerated from rest in the outer low pressure region, where the field is inhomogeneous, do not reach the highest energies, whilst those injected in the region of high pressure are fully accelerated in a field sufficiently homogeneous for the purpose.

The weight of any objection to the theory on the grounds of inhomogeneity of the field is reduced when it is considered firstly that the orbits do not loop the origin, and secondly that most particles which attain final energies in excess of 10^9 ev always move on a path with a radius of curvature less than about 0·01 times the radius of the sunspot. Consequently, they will spend the acceleration time in a region of field where the lines are closely parallel, even though a complete drawing of the field (Hoyle (1949 b)) may suggest strong overall inhomogeneity. A calculation of the particle orbits in a non-uniform field would be most difficult, and hardly appears worth while pending more information on the nature of the field, and also the spatial variation of pressure near the sun's surface, which affects the injection energy.

Only those particles that are released at the instant a spot comes into or recedes from view as the sun rotates may travel directly to the earth, and these could be responsible for the increase in cosmic ray intensity sometimes observed during periods of intense solar activity (Forbush, Gill and Vallarta (1949)). But this must be considered doubtful on the basis of evidence concerning the latitude effect of such increases, summarized by Elliot (Wilson (1951 b)) in a recent book on progress in cosmic ray physics. This suggests that the rays concerned have energies of only a few times 10^9 ev, and are produced by some process other than that responsible for the normal cosmic radiation. Those particles that do not travel directly to the earth (or other planets) are trapped in the Richtmeyer–Teller field and circulate for about 10^6 years before striking the earth. If the sun is the main source of cosmic rays, the existence of this extensive magnetic field of strength about 10^{-5} gauss is necessary to smooth out the observed cosmic ray intensity in both space and time. Since the solar magnetic field is now considered (Alfven (1951), Thiessen (1952)) to be very

much weaker than was supposed some years ago, it is unlikely that its presence will complicate either the acceleration or extraction process. In any event most spots occur near the sun's equator, so that the component of the dipole field normal to the spot will be too small to affect the orbits even in the early stages of acceleration. Were the sun's magnetic field of order 50 gauss, as once supposed, particles of energy less than 10^{10} ev could escape from the sun only at high latitudes, although Forbush *et al.* (1949) did show how suitable combinations of this field with the field of the spot produce tunnels through which particles of energy as low as 10^{9} ev may escape.

Electrons also will be accelerated in the spot whilst its field is changing, but will not attain energies in excess of about 10^{9} ev because of loss of energy by radiation. The ionization loss curves for electrons have maxima at energies of about 100 ev. For protons and heavier particles radiative energy losses are negligible. The electron radiation may be related to the 'disturbed' solar noise observed during periods of sunspot activity (Twiss (1952)). Secondary mesons and neutrons will degenerate before reaching the earth.

§7. Comparison with other Theories, and Conclusions

Various processes which would accelerate charged particles to cosmic ray energies have been proposed. The most successful is that of Fermi (1949) and Chang-Yun Fan (1951), which permits very high energies and also leads quite naturally to an inverse power law for the energy spectrum. The process involves acceleration by electromagnetic induction when charged particles collide with the magnetic fields associated with streams of ionized matter passing through interstellar space. Several mechanisms involving solar phenomena which would produce particles of energy of order 10^{8}—10^{10} ev have been recently described by Alfven (1946 and 1949), Bagge and Biermann (1948), Menzel and Salisbury (1948), Kiepenheuer (1950) and Terletski (1949). A brief description of some of them has been given by Elliot (Wilson (1951 c)). Theories involving acceleration in straight static or increasing electric fields, as distinct from induced fields of the type described in this paper, must be doubted for the reason that the spectrum varies at most only logarithmically with energy. However, as already mentioned, there exist solar phenomena which produce particles of energy at most a few times 10^{9} ev, whose number-energy spectrum is as yet unknown. These leave the sun in relatively intense beams, parts of which often proceed more or less directly to the earth.

The present theory has several advantages over Fermi's :

(*a*) The injection energy required is of order 1 mev, whereas for Fermi's process it is about 1 000 mev. Fermi presupposes the existence of an injection mechanism such as the process described in this paper.

(*b*) The index in the power law of Fermi's spectrum depends linearly on the distance between magnetic clouds, their velocity, and the average density of matter in interstellar space. Recently Unsold (1951) has concluded that the value of that index is about 200. For the present case

it cannot lie outside the range 1·7 to 3·0 independent of what sunspot parameters are assumed.

(*c*) Experimental information about sunspot magnetic fields is more certain than that about magnetic clouds. We are not concerned with the mechanism producing the sunspot field, only with the fact that it exists.

An objection raised by Saha (1951) to a solar origin for the cosmic rays is that, although the sun exhibits no neon spectrum, neon nuclei are observed among the primary rays. Thus it may be that the results of the idealized theory presented here are fortuitous, although there is no reason to suppose that spots do not occur on other stars.

Acknowledgments

We are grateful to several members of the Physics and Mathematical Physics Departments for informative discussions, particularly Dr. R. H. Dalitz, who also pointed out that eqn. (3) is a solution of the equations of motion, and Dr. Martin Johnson. We thank Professor P. M. S. Blackett of the University of Manchester for bringing the papers of Terletski and Bagge and Biermann to our notice, and for a most constructive discussion which assisted the writing of this paper.

References

Alfven, H., 1946, *Nature, Lond.*, **158,** 618; 1949, *Phys. Rev.*, **75,** 1732; 1950, *Cosmical Electrodynamics*, p. 230 (Oxford University Press); 1951, *Nature, Lond.*, **168,** 1036.

Bagge, E., and Biermann, J., 1948, *Naturwiss.*, **35,** 120.

Blackett, P. M. S., 1950, *Report on the Bombay Conference.*

Blanch, G., Lowan, A. N., Marshak, R. E., and Bethe, H. A., 1941, *Astrophys. J.*, **94,** 37.

Bradt, H. L., and Peters, B., 1950, *Phys. Rev.*, **77,** 54.

Chang-Yun Fan, 1951, *Phys. Rev.*, **82,** 211.

Fermi, E., 1949, *Phys. Rev.*, **75,** 1169.

Forbush, S. E., Gill, P. S., and Vallarta, M. S., 1949, *Rev. Mod. Phys.*, **21,** 44.

Hilberry, N., 1941, *Phys. Rev.*, **60,** 7.

Hoyle, F., 1949 a, *Recent Researches in Solar Physics*, p. 2. (Cambridge University Press) ; 1949 b, *Ibid.*, p. 3.

Kerst, D. W., and Serber, R. B., 1941, *Phys. Rev.*, **60,** 53.

Kiepenheuer, K. O., 1950, *Phys. Rev.*, **78,** 809.

Lemaitre, G., and Vallarta, M. S., 1936, *Phys. Rev.*, **50,** 493.

Menzel, R. H., and Salisbury, W. W., 1948, *Nucleonics*, **2,** 67.

Richtmeyer, R. D., and Teller, E., 1949, *Phys. Rev.*, **75,** 1729.

Saha, M. N., 1920, *Phil. Mag.*, **40,** 472 ; 1951, *Nature, Lond.*, **167,** 476.

Swann, W. F. G., 1933, *Phys. Rev.*, **43,** 217.

Terletski, P., 1949, *Journ. Exp. and Theor. Phys.* (*U.S.S.R.*), **19,** 1059.

Thiessen, G., 1952, *Nature, Lond.*, **169,** 147.

Twiss, R. Q., 1952, *Nature, Lond.*, **169,** 185.

Unsold, A., 1951, *Phys. Rev.*, **82,** 857.

Van Allen, J. A., and Singer, S. F., 1950, *Phys. Rev.*, **78,** 819.

Wilson, J. G., 1952 a, *Progress in Cosmic Ray Physics*, p. 300 (North Holland Publishing Co., Amsterdam) ; 1952 b, *Ibid.*, p. 502; 1952 c, *Ibid.*, p. 509.

Winckler, J. R., Stix, T., Dwight, K., and Sabin, R., 1950, *Phys. Rev.*, **79,** 656.

XLI. *The Photodisintegration of the Deuteron, and the Measurement of γ-Ray Energies with Photographic Emulsions containing Heavy Water*

By W. M. Gibson
H.H. Wills Physical Laboratory, University of Bristol
and
T. Grotdal, J. J. Orlin and B. Trumpy
Fysisk Institutt, University of Bergen*

[Received January 22, 1952]

Summary

Photographic plates soaked in heavy water have been used in a study of the photodisintegration of deuterons by γ-rays of energy 6·1 and 7·0 Mev. The principles of the method were described in an earlier paper, but the technique has been improved and the accuracy of the results increased.

The method has been shown to be valuable for the measurement of γ-ray energies in the interval 4–12 Mev. The accuracy obtained at 6 Mev is about 40 Kev, while γ-ray lines with an energy difference of 0·4 Mev could be resolved.

The angular distribution of the protons has been measured, and the values $a/b=0{\cdot}05\begin{cases}+0{\cdot}03\\-0{\cdot}02\end{cases}$ and $k=0{\cdot}02\pm0{\cdot}08$ have been obtained for the constants in the equation

$$\sigma(\phi)=a+b\sin^2\phi(1+k\cos\phi).$$

The value of a/b is lower than that found by Goldhaber, but no exact theoretical values are available for comparison.

§1. Introduction

Although the photodisintegration of the deuteron has been very widely studied, both experimentally and theoretically, there are still considerable uncertainties in both observed and predicted angular distributions of the emitted protons (e.g. Goldhaber 1950).

A discussion of some of the methods used in the measurement of this angular distribution, and of the difficulties to be overcome, has been given in an earlier paper (Gibson, Grotdal, Orlin and Trumpy 1951, henceforth referred to as A). The greater part of paper A was, however, devoted to an extension of the method used by Goldhaber (1951), Waffler and Younis (1949), Hough (1950) and others, in which photographic plates soaked with heavy water are exposed to γ-radiation. Descriptions

* Communicated by Professor C. F. Powell, F.R.S.

were given of the methods of exposing the plates to the 6 and 7 MeV γ-rays from the bombardment of fluorine with protons, examining the proton tracks produced in the emulsion by photodisintegration, calculating the energy of the γ-ray responsible for each track, and obtaining the angular distribution of the protons. It was pointed out that improvement of the technique would give it considerable value as a means of measuring γ-ray energies, and that further measurements should give useful information about the angular distribution.

Exposures have now been made under more satisfactory conditions, and a much greater number of tracks has been examined. In this paper the method of measuring γ-ray energies will be given in a form which will, it is hoped, prove convenient for others who wish to use it; the results for the angular distribution of the protons will be presented and discussed.

§2. Technique

When emulsions soaked in heavy water are being used, the amount by which the emulsion swells must be known, and should vary as little as possible during the exposure. The way in which the thickness of Ilford C2 emulsions increases with time after immersion was therefore studied by direct micrometer measurements. Table 1 shows the thickness, at different times, of an emulsion which was 212·5 μ thick before immersion.

Table 1

Time in hours	0	0·5	1·0	1·5	2·5	3·5	4·5	7·5	10·5	24·0	49·0
Thickness in μ	212·5	535	617·5	649	674·5	686	696·5	712·5	720	730	747·5

After 24 hours, the thickness is increasing at the rate of about 0·1% per hour; a soaking time of 21 hours was chosen, since the subsequent swelling would not reach 1% in an exposure lasting less than 10 hours.

As in the earlier work, the plates were exposed to the γ-rays produced by bombarding a thick target of calcium fluoride with 1·4 MeV protons from the van de Graaff generator at the University of Bergen. However, the plates were placed closer to the target, some of them so that their centre portions were 5 cm away and received the γ-rays perpendicularly, others so that γ-rays entered the regions 5 cm from the target at a glancing angle of about 18°. The plates were held in boxes so designed that there was nothing to restrict the swelling of the emulsions although the plates were held rigidly.

It has been calculated that the range-energy relation for protons in an Ilford C2 emulsion which has absorbed enough heavy water to increase its thickness by a factor C, should be given by:

$$R=(15{\cdot}15+1{\cdot}10\,C)E^{1{\cdot}65}$$

if R is the range in microns, and E the energy in MeV. This relation should be sufficiently accurate for values of C between 3 and 5 and for

energies from 1·5 to 3 MeV, but slight changes in the numerical constants would make it applicable to other values of C and E. For the emulsions used in the work described here, the value of C was 3·94.

The above range-energy relation has been obtained from the data of Livingston and Bethe (1937), Bethe (1950) and Rotblat (1951); the differential stopping powers of D_2O and the dry emulsion, assumed to mix without any chemical reaction which would cause an overall change in density, gave the differential stopping power of the wet emulsion, and hence the relation between the energy of a proton and its range. In the wet emulsion, however, the silver bromide grains are comparatively widely spaced, and the observed range, from the centre of the first grain in the track to the centre of the last grain, is less than the true range. The average difference between the true and measured ranges is, under the conditions of this experiment, approximately equal to the difference between the average grain spacing in an emulsion exposed while wet, and that in an ordinary dry emulsion. Measurements showed that this was 1·25μ, and this correction is included in the range-energy relation given.

It was found that emulsions soaked in ordinary water swelled more than those soaked in heavy water. This was probably due to a difference in pH between the liquids used, rather than to a difference in the properties of heavy and ordinary water. But whatever the reason for the effect, allowance must be made for it if an emulsion soaked in H_2O is to be used to show how many tracks are due to processes other than the photodisintegration of D.

It is necessary to know not only the value of C, but also the shrinkage factor, i.e. the ratio of the thickness of the emulsion during exposure to that after processing. Direct measurements showed that the value of this factor was between 9·5 and 11·0, varying with the humidity of the air in which the processed plate was kept. In most of the work described the thickness of the processed plate was measured daily, and an appropriate shrinkage factor used in the subsequent calculations.

The shrinkage factor is so large under these conditions that the angle of dip of a track in the wet emulsion, as calculated from depth measurements after processing, is very inaccurate. The accuracy of the measurements of angle is limited by that of the depth measurements, and the greatest care was taken to make the errors as small as possible. In any future work, however, an attempt would be made to increase the thickness of the emulsion after processing, e.g. by the technique suggested by Danysz and Yekutieli (1951). It appears to be possible to impregnate emulsions with resin, so that their thickness when dry is comparable with that before processing. The upper limit to the thickness is set only by the impregnation technique, but even if it could be increased by no more than a factor of 2, the errors in the angles of dip would be halved; there would be the additional advantage that an emulsion which was half resin would be affected very little by atmospheric humidity, and a constant shrinkage factor could be used.

§3. Measurement of γ-ray Energies

When γ-ray energies are to be calculated, plates exposed at right angles to the incident radiation contain more tracks suitable for accurate measurement. The greatest accuracy is obtained by imposing a strict limit on the angle of dip, though of course, this increases the area of plate which must be searched for a given number of tracks.

In the series of measurements directed towards obtaining the greatest possible accuracy and resolution, 593 tracks, each with an angle of dip less than 45°, were examined ; the horizontal and vertical projections of the range of each track were measured, and the range and angle of dip of the track when formed were calculated from these and the appropriate shrinkage factor.

If, in the notation of paper A, θ is the angle between the γ-ray and the proton, and δ is the angle of dip of the proton,

$$\theta = 90° - \delta$$

since the measurements were restricted to a small area of plate over the whole of which the angle of incidence of the γ-rays was effectively 90°. It follows from equations given in paper A that the γ-ray energy is

$$\begin{aligned} E_\gamma &= 2{\cdot}20 + 2Ep(1 - 0{\cdot}102 \cos\theta) \\ &= 2{\cdot}20 + 2Ep(1 - 0{\cdot}102 \sin\delta). \end{aligned}$$

The energy Ep and the range R of a proton are related by the equation :

$$Ep^{1{\cdot}65} = R/(15{\cdot}15 + 1{\cdot}10\,C),$$

therefore E_γ is given by :

$$E_\gamma = 2{\cdot}20 + 2(1 - 0{\cdot}102 \sin\delta)[R/(15{\cdot}15 + 1{\cdot}10\,C)]^{0{\cdot}606}.$$

Several types of nomogram for reducing the labour of the calculation of $E\gamma$ from each track were tried, but they were not very satisfactory because a slight error in drawing a line of constant $E\gamma$ led to a relatively greater percentage error in the distance between adjacent lines, and hence in the number of tracks allocated to each interval of $E\gamma$. The value of $R(1 - 0{\cdot}102 \sin\delta)^{1{\cdot}65}$ was therefore obtained for each track ; the limits of this quantity for successive 0·1 Mev intervals of $E\gamma$ were calculated and used to give the number of tracks within each interval of $E\gamma$.

The γ-ray energy distribution obtained from 443 tracks with $|\delta| \leqslant 30°$ is given in fig. 1. It shows clearly the presence of two groups of γ-rays of mean energy 6·15 and 7·05 Mev. The accuracy of these figures is determined by the following factors :

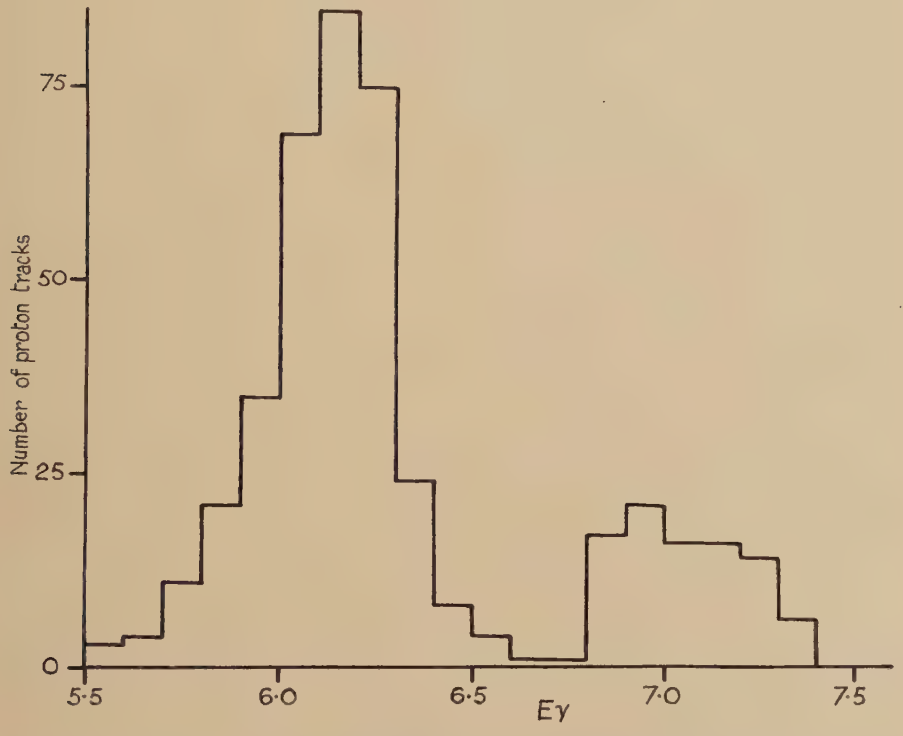
1. The range-energy relation : this may contain errors of 1%, which would lead to errors of 0·025 Mev in the values of $E\gamma$.

2. The value of C, the factor by which the emulsion swells : this was obtained by micrometer measurements on emulsions from the same batch as those exposed. The consistency of the results suggests that the probable error of C is about 0·1 ; this would have affected the range of a particle of given energy by 0·5% and caused a further error of 0·012 Mev in $E\gamma$.

3. The spread of the calculated values of $E\gamma$: the mean deviation of the 336 individual values grouped around 6·15 MEV is 0·15 MEV, and the probable error of the mean should therefore be only 0·01 MEV. The probable error of the mean energy 7·05 MEV, based on 60 tracks, is 0·03 MEV.

4. Correction for dip : the measured angle of dip of an individual track may be in error by as much as 5°. This would lead to an error of only 0·1% in the factor $(1-0{\cdot}102 \sin\delta)^{1{\cdot}65}$, but could give a 5% error in the calculated range. Only systematic errors in δ will affect the mean value of $E\gamma$, and the probable magnitude of their effects is estimated as 0·025 MEV.

Fig. 1



5. Systematic errors in measurement of range : it is difficult to estimate the effect of the occasional inclusion of a background grain at one end of the track, or omission of a grain which should, in fact, have been included. These errors would not be expected, on the average, to amount to 1% of the range, and a provisional value of 0·7% may be set for the probable error.

The probable errors of the values 6·15 MEV and 7·05 MEV are therefore estimated as 0·04 MEV and 0·05 MEV respectively. The work of Burcham and Freeman (1949) and Chao, Tollestrup, Fowler and Lauritsen (1950) has shown that bombardment of fluorine with protons should give rise to γ-ray lines of energy 6·14, 6·91 and 7·11 MEV ; the last two, if not resolved, should appear as a composite group of mean energy 7·01 MEV,

with protons of energy 1·4 MeV. The observed values (6·15±0·04) MeV and (7·05±0·05) MeV, are in good agreement with the more accurate values 6·14 and 7·01 MeV obtained from measurements on the associated α-particles.

The resolution shown in fig. 1 is considerably greater than those described in paper A and by Goldhaber (1951). With this resolution, two γ-ray lines of similar intensity 0·4 MeV apart would give a distribution showing two peaks, while if the energy difference were 0·7 MeV there would be almost complete separation of the groups.

For certain types of work, this technique for measuring γ-ray energies should be very useful. Its principal advantage is that the operations to be conducted during the exposure are very simple. The only background effect is from neutrons, and this can easily be estimated by exposing a plate soaked in H_2O at the same time. The accuracy of the γ-ray energy measurements has been shown to be as high as that given by other methods (see, for instance, Rasmussen, Hornyak, Lauritsen and Lauritsen 1950), while that of intensity measurements is limited only by the accuracy with which the cross section of the deuteron for photodisintegration is known (Wilkinson 1950), and by the number of tracks measured.

Fairly weak sources can be used : the work described here was based on plates containing about 10 useful tracks per mm², and exposed to the γ-rays from a calcium fluoride target bombarded by protons carrying a total charge of 18 millicoulombs ; since a plate can be exposed for 10 hours without appreciable fading of the latent image, and proton beams of up to 200 microamperes are now available, 7 200 millicoulombs could be used. With many targets the γ- and X-rays in such an exposure would fog the emulsion, but the experiments of Gibson, Green and Livesey (1947) suggest that one coulomb could often be used, with the plates 5 cm from the target. It would be possible, though less satisfactory, to work with plates containing only 4 useful tracks per mm², and therefore it seems reasonable to claim that reactions yielding γ-ray lines 100 times less intense than those from the reaction $F(p, \alpha\gamma)O^{16}$ could be studied without undue difficulty. The weakest source of γ-rays which could be used would emit about 3×10^6 quanta per second.

The technique is satisfactory, however, only between about 4 and 12 MeV ; the threshold for photodisintegration of the deuteron is 2·2 MeV, and γ-rays of energy less than 4 MeV would not give tracks long enough for accurate measurement. At energies above 12 MeV, the number of protons from (γ, p) reactions in silver and bromine becomes appreciable, and measurements have to be made with a large background of such tracks.

§4. Angular Distribution of the Protons

Over 5 000 proton tracks have been observed and measured in plates exposed so that the γ-rays entered obliquely. For each track, the horizontal and vertical projections of the range, and the direction projected

onto a horizontal plane, have been measured, and the true direction and range calculated. The angle θ between the directions of the γ-ray and the proton was obtained by a method of stereographic projection (see paper A).

Some tracks were omitted because the direction of motion of the proton was uncertain ; it was confirmed that this did not influence the relative numbers allocated to the different intervals of δ and ϕ.

For the greater proportion of the tracks, R and θ were plotted on a polar diagram, and the γ-ray energy and centre-of-mass angle ϕ were found from a map which could be superimposed on this diagram (see paper A). Towards the end of the work, however, it was decided that there was no need to obtain the angular distributions of protons produced by the 6 and 7 MeV γ-rays separately. The equation :

$$\tan\theta=\sin\phi/(\cos\phi+0{\cdot}051)$$

was therefore used to allocate each proton track to the appropriate 10° interval of ϕ, the γ-ray energy not being calculated ; it was confirmed, however, that the range of each proton lay within the required limits, since there was a small number of tracks which must have been due to other processes. Measurements in a plate exposed while soaked with H_2O showed that the number of spurious tracks likely to have been included was negligible.

The tracks were also divided into four groups, according to whether their angle of dip δ lay in the intervals 0–22·5°, 22·5°–45°, 45°–67·5°, or 67·5°–90°. It was found that for all except one set of measurements, the ratios of the total numbers of tracks in the first three intervals of δ were close to those expected. There were other reasons for supposing that this set was unreliable, and it was rejected. Tracks in the fourth interval were quite unreliable on account of the inaccuracy of the vertical measurements, and the calculation of the angular distribution was based only on the two most reliable intervals, 0–22·5° and 22·5°–45°. For these two intervals, it was calculated that the apparent angular distribution was affected to a negligible extent by the escape of protons from the surfaces of the emulsion.

The method of obtaining the true angular distribution from the data available—namely the number of tracks found in each 10° interval of ϕ and each 22·5° interval of δ—is outlined in the appendix. It is laborious, but other simpler methods involve approximations which, in certain cases, can mask the effects that are being sought.

The values of a/b and k, when the differential cross-section is expressed in the form

$$\sigma(\phi)=a+b\sin^2\phi(1+k\cos\phi)$$

are given in table 2.

The effect of tracks from processes other than the photodisintegration of D, and of mistakes in measurement or calculation, would be to increase the apparent value of a/b. If the number of these spurious tracks is taken

to be 1·5% of the total, and they are assumed to be distributed isotropically, the apparent value of a/b is increased by 0·01. The lower limits of uncertainty given for the values of a/b in table 2 have therefore been set 0·01 below those given by the method outlined in the Appendix.

Table 2

δ limits	Place of measurement	No. of tracks	a/b	k
0–22·5°	Bergen	1024	$0{\cdot}07\,{}^{+0{\cdot}11}_{-0{\cdot}05}$	$0{\cdot}04\pm0{\cdot}14$
0–45°	Bergen	2099	$0{\cdot}04\,{}^{+0{\cdot}06}_{-0{\cdot}025}$	$0{\cdot}06\pm0{\cdot}12$
0–45°	Bristol	934	$0{\cdot}06\,{}^{+0{\cdot}10}_{-0{\cdot}03}$	$-0{\cdot}06\pm0{\cdot}18$
0–45°	Total	3033	$0{\cdot}05\,{}^{+0{\cdot}03}_{-0{\cdot}02}$	$0{\cdot}02\pm0{\cdot}08$

The different values for a/b agree very well, and the final value of $0{\cdot}05\,{}^{+0{\cdot}03}_{-0{\cdot}02}$ is of interest since it is appreciably lower than that obtained by Goldhaber (1951) ($0{\cdot}15\pm0{\cdot}06$). This difference is greater than can be accounted for by the fact that the mean energy of the γ-rays used by Goldhaber was 6·8 MeV, while that for the present work was 6·3 MeV.

The contribution to a/b from the magnetic transitions is expected to be 0·02 (Hansson and Hulthen 1949); it appears that the contribution from interference between the final ${}^3p_j$ states is likely to be of the same order, and not large enough to account for Goldhaber's result, although exact calculations have not been made. If the contribution from the magnetic transitions is assumed to be 0·02, the measured value of the contribution from interference is seen to be $0{\cdot}03\,{}^{+0{\cdot}03}_{-0{\cdot}02}$. The comparison of this, and more exact values which may be obtained in the future, with those given by theoretical calculations, will be of great interest.

The final value of k is $0{\cdot}02\pm0{\cdot}08$. This is lower than the value 0·14 predicted by Marshall and Guth (1949, 1950), but the difference is probably not significant.

Acknowledgments

The authors' thanks are due to Misses J. A. Burnell, B. M. Pillinger and M. J. Chapple, and Messrs. K. F. Kordt, P. Johannessen and S. Skaale, for the microscope work, to Dr. E. J. Burge and Mr. G. André for assistance with the calculations, and to Professor C. F. Powell for his advice and encouragement.

APPENDIX

Calculation of Angular Distribution

The cross section per unit solid angle, in the centre of mass system, is of the form

$$a+b\sin^2\phi(1+k\cos\phi)$$

and the problem is to find the best values of a/b and k.

The cross section per unit interval of ϕ is proportional to :

$$\{a+b\sin^2\phi(1+k\cos\phi)\}\sin\phi$$

The cross section for an interval ϕ_1 to ϕ_2 is therefore proportional to :

$$a\int_{\varphi_1}^{\varphi_2}\sin\phi\,d\phi+b\int_{\varphi_1}^{\varphi_2}\sin^3\phi\,d\phi+bk\int_{\varphi_1}^{\varphi_2}\sin^3\phi\cos\phi\,d\phi.$$

But if certain conditions are imposed on the value of the angle of dip δ only a fraction I (a function of ϕ) of the tracks at each ϕ will be included ; the observed number of tracks within each interval ϕ_1 to ϕ_2 ought therefore to be equal to

$$a\int_{\varphi_1}^{\varphi_2}I\sin\phi\,d\phi+b\int_{\varphi_1}^{\varphi_2}I\sin^3\phi\,d\phi+bk\int_{\varphi_1}^{\varphi_2}I\sin^3\phi\cos\phi\,d\phi.$$

I is calculated as a function of ϕ for each of the two intervals of δ, 0–22·5° and 0–45°.

The products $I\sin\phi$, $I\sin^3\phi$ and $I\sin^3\phi\cos\phi$ are plotted against ϕ, and the values of

$$\int_{\varphi_1}^{\varphi_2}I\sin\phi\,d\phi=X,\quad\int_{\varphi_1}^{\varphi_2}I\sin^3\phi\,d\phi=Y\text{ and }\int_{\varphi_1}^{\varphi_2}I\sin^3\phi\cos\phi\,d\phi=Z$$

are obtained for each 10° interval of ϕ, by numerical integration.

This procedure gives, for each of the intervals of dip, a series of values of X, Y and Z such that the observed number of tracks (N) in a given 10° interval of ϕ should be $aX+bY+bkZ$.

In fact, it is necessary to find the best values of a, b and bk by the method of least squares. The equations to be solved are :

$$\begin{aligned}aX_1+bY_1+bkZ_1&=N_1\\aX_2+bY_2+bkZ_2&=N_2\\\vdots\quad&\quad\vdots\quad\vdots\quad\vdots\end{aligned}$$

Given equal weight these become :

$$\begin{aligned}aX_1/\sqrt{N_1}+bY_1/\sqrt{N_1}+bkZ_1/\sqrt{N_1}&=\sqrt{N_1}\\aX_2/\sqrt{N_2}+bY_2/\sqrt{N_2}+bkZ_2/\sqrt{N_2}&=\sqrt{N_2}\\\vdots\quad&\quad\vdots\quad\vdots\quad\vdots\end{aligned}$$

The ' normal equation for a ' is therefore :

$$a\Sigma(X^2/N)+b\Sigma(XY/N)+bk\Sigma(XZ/N)=\Sigma X.$$

Similarly, the other normal equations are :

$$a\Sigma(XY/N)+b\Sigma(Y^2/N)+bk\Sigma(YZ/N)=\Sigma Y,$$

$$a\Sigma(XZ/N)+b\Sigma(YZ/N)+bk\Sigma(Z^2/N)=\Sigma Z.$$

Simultaneous solution of these three gives the best values of a, b and bk, and hence of a/b and k.

This is done for each of the two dip intervals. When no track is observed in a particular interval of ϕ, the appropriate terms in the sums $\Sigma(X^2/N)$, etc., are replaced by X^2, etc., and the appropriate terms in the sums ΣX, ΣY and ΣZ are omitted ; this gives the correct result, and assumes that a measurement of zero has the same weight as a measurement of 1.

The probable error of any result can be found by the χ^2 method (Fisher 1946). In this case, a value of $\chi^2=15{\cdot}5$ corresponds to a probability 0·50, and the probable errors quoted in this paper are those which give $\chi^2=15{\cdot}5$.

References

Bethe, 1950, *Rev. Mod. Phys.*, **22,** 213.
Burcham and Freeman, 1949, *Phys. Rev.*, **75,** 1756.
Chao, Tollestrup, Fowler and Lauritsen, 1950, *Phys. Rev.*, **79,** 108.
Danysz and Yekutieli, 1951, *Phil. Mag.*, **42,** 1185.
Fisher, 1946, *Statistical Methods for Research Workers.* Oliver and Boyd, Edinburgh.
Gibson, Green and Livesey, 1947, *Nature, Lond.*, **160,** 534.
Gibson, Grotdal, Orlin and Trumpy, 1951, *Phil. Mag.*, **42,** 555.
Goldhaber, 1951, *Phys. Rev.*, **81,** 930.
Hansson and Hulthen, 1949, *Phys. Rev.*, **76,** 1163.
Hough, 1950, *Phys. Rev.*, **80,** 1069.
Livingston and Bethe, 1937, *Rev. Mod. Phys.*, **9,** 245.
Marshall and Guth, 1949, *Phys. Rev.*, **76,** 1879, 1880 ; 1950, *Ibid.*, **78,** 738.
Rasmussen, Hornyak, Lauritsen and Lauritsen, 1950, *Phys. Rev.*, **77,** 617.
Rotblat, 1951, *Nature, Lond.*, **167,** 550.
Waffler and Younis, 1949, *Helv. Phys. Acta*, **22,** 414.
Wilkinson, 1950, *Harwell Conference Report.*

XLII. *Electron-Electron Interaction in Heavy Atoms*

By G. E. Brown
Department of Mathematical Physics, University of Birmingham*

[Received January 4, 1952]

§ 1. Introduction

The derivation from quantum electrodynamics of a Hamiltonian describing the interaction of two fast-moving electrons in a heavy atom to order $\alpha=e^2/4\pi\hbar c$ will be reported here. The case considered is exemplified by the interaction of two K-electrons in a heavy element, where v/c, the ratio of the electron velocity to that of light, is of order $Z\alpha$. This case is essentially different from the situation in light atoms where v/c and the radiative reaction are both small and where one can expand in both (Breit 1929, Oppenheimer 1930). In heavy atoms, one cannot expand in v/c, but it is still possible to expand in the radiative interaction, or in powers of α. The large electron-nuclear interaction should be taken into account exactly, however.

For comparison, the electron-electron interaction Hamiltonian of Breit, obtained by an expansion in v/c, which describes the interaction of two slow-moving electrons, will be given here :

$$H_{\text{int}}^{(B)}=\frac{e^2}{4\pi}\left[\frac{1}{|\mathbf{r}-\mathbf{r}^1|}-\frac{\boldsymbol{\alpha}\cdot\boldsymbol{\alpha}^1}{2|\mathbf{r}-\mathbf{r}^1|}-\frac{\boldsymbol{\alpha}\cdot(\mathbf{r}-\mathbf{r}^1)\,\boldsymbol{\alpha}^1\cdot(\mathbf{r}-\mathbf{r}^1)}{2|\mathbf{r}-\mathbf{r}^1|^3}\right] \quad . \quad (1)$$

where the $\boldsymbol{\alpha}$'s are Dirac's operators for $\mathbf{v}/c$, $\boldsymbol{\alpha}$, $\mathbf{r}$ refer to the one electron and $\boldsymbol{\alpha}^1$, $\mathbf{r}^1$ to the other. The first term in the square brackets is the Coulomb term. Two times the second term is the magnetic interaction, and the remainder—the third term minus the second—represents the retardation of the Coulomb potential. The classical interaction for two slow-moving electrons can be obtained from (1) by replacing $\boldsymbol{\alpha}$, $\boldsymbol{\alpha}^1$ by $\mathbf{v}$, $\mathbf{v}^1$.

It will be shown that the interaction term, to order α, which should be employed for electrons in heavy atoms is

$$H_{\text{int}}=\frac{e^2}{4\pi}\frac{(1-\boldsymbol{\alpha}\cdot\boldsymbol{\alpha}^1)\exp\left[i|E_m-E_n|(|\mathbf{r}-\mathbf{r}^1|)\right]}{|\mathbf{r}-\mathbf{r}^1|} \quad . \quad . \quad . \quad (2)$$

where E_n and E_m are the energies of the initial and final hydrogen-like states which are eigenfunctions of non-interacting electrons in the nuclear potential. In cases where $E_m=E_n$; e.g., direct diagonal matrix elements, (2) takes the simple form :

$$H_{\text{int}}=\frac{e^2}{4\pi}\frac{1-\boldsymbol{\alpha}\cdot\boldsymbol{\alpha}^1}{|\mathbf{r}-\mathbf{r}^1|} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2.1)$$

and it is seen that the retardation of the Coulomb potential occurring in (1) is entirely absent here.

---

* Communicated by Professor R. E. Peierls.

§2. Derivation

The Schrödinger equation for the wave function describing the electron field, the radiation field and the interaction between these two fields, may be written*

$$i\frac{\partial\Psi(t)}{\partial t}=\{H_0+H_e+H_i\}\Psi(t) \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

where H_0 is the Hamiltonian for the free electron field plus uncoupled radiation field, H_e describes the interaction of the electrons with an external field, here taken to originate from a stationary point nucleus of charge Ze situated at the coordinate origin ; i.e.

$$H_e=\frac{Ze^2}{4\pi}\int d^3r\bar{\psi}(\mathbf{r})\gamma_4\psi(\mathbf{r})\frac{1}{r}, \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

the $\bar{\psi}$ and ψ are the electron field operators, to be explained below, and $\mathbf{r}$ is the space coordinate associated with the electron field. Finally,

$$H_i=-\int d^3rj_\mu(\mathbf{r})A_\mu(\mathbf{r}), \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

with $j_\mu(\mathbf{r})=-ie\bar{\psi}(\mathbf{r})\gamma_\mu\psi(\mathbf{r})$ and $A_\mu(\mathbf{r})$ the radiation field operator. Now the philosophy here is to include H_e in the zero order Hamiltonian. As a consequence, instead of making the usual transformation of Schwinger (1949) removing H_0 from (3) and giving the field operators in H_e and H_i the time dependence of free-field operators, one should make the transformation defined by

$$\Psi'(t)=\exp[i(H_0+H_e)t]\Psi(t) \; ; \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

then

$$i\frac{\partial\Psi'(t)}{\partial t}=H_i(t)\Psi'(t), \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (7)$$

where

$$H_i(t)=\exp[i(H_0+H_e)t] \quad H_i\exp[-i(H_0+H_e)t]. \quad . \quad . \quad (8)$$

The $\bar{\psi}(\mathbf{r})$ and $\psi(\mathbf{r})$ in H_i can be resolved in various ways into a sum over one-electron eigenfunctions. Here it is obviously convenient to pick the hydrogen-like eigenfunctions of the Dirac equation for one electron moving in a potential Ze/r. This is the equation in configuration space which corresponds to the zero order equation here. In terms of these

$$\psi(\mathbf{r})=\sum_n b_n\psi_{Hn}(\mathbf{r}). \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (9)$$

Here n runs through all eigenstates, both discrete and continuous, and b_n is an operator which annihilates an electron in the nth state if $E_n>0$ or creates a positron if $E_n<0$. Similarly,

$$\bar{\psi}(\mathbf{r})=\sum_m b_m{}^*\bar{\psi}_{Hm}(\mathbf{r}) \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (10)$$

---

* Natural units ($\hbar=c=1$) will be employed. The $\Psi(t)$ is Schwinger's $\Psi(\sigma)$ with the choice of σ as a constant time plane. Otherwise the notation is the same as in Schwinger (1948).

and b_m^* creates an electron ($E_m>0$) or annihilates a positron ($E_m<0$). Under the transformation (6)–(8),

$$\psi_{Hn}(\mathbf{r}) \rightarrow \psi_{Hn}(\mathbf{r}, t)=\psi_{Hn}(\mathbf{r}) \exp(-iE_n t).$$

The first order effects of $H_i(t)$ in (7) are eliminated by means of the transformation

$$\exp[iS(t)]=\exp\left[(-i/2)\int_{-\infty}^{\infty} dt^1 \epsilon(t-t^1)H_i(t^1)\right]; \quad . \quad . \quad . \quad (11)$$

if $\Psi''(t)$ is defined by $\Psi'(t)=\exp[iS(t)]\Psi''(t)$, then

$$i\frac{\partial \Psi''(t)}{\partial t} = \{\tfrac{1}{2}i[S(t), H_i(t)]-\tfrac{1}{3}[S(t), [S(t), H_i(t)]]+ \ldots\}\Psi''(t). \quad . \quad . \quad (12)$$

Only the first commutator will be considered, since it represents all effects to within order e^2.

The 'one-electron parts'—in the terminology of Schwinger—of $H_T(t) \equiv \frac{1}{2}i[S(t), H_i(t)]$ give not only the infinite lowest order mass renormalization, but also Lamb-shift terms for a single electron in potential Ze/r. (The one-electron parts give the self-energies of electrons in hydrogen-like states and these self-energies differ from free-electron self-energies by Lamb-shift terms.) These one-electron parts can be separated off unambiguously from the two-electron parts, which are of interest here. One has

$$H_T(t)=\frac{e^2}{2}\int d^3r^1\, dt^1\, d^3r\, \bar{\psi}(\mathbf{r}^1, t^1)\gamma_\mu\psi(\mathbf{r}^1, t^1) \times \overline{D}(\mathbf{r}^1-\mathbf{r}, t^1-t)\bar{\psi}(\mathbf{r}, t)\gamma_\mu\psi(\mathbf{r}, t) \quad . \quad . \quad . \quad (13)$$

where $\overline{D}(\mathbf{r}^1-\mathbf{r}, t^1-t)=-\frac{1}{2}\epsilon(t^1-t)D(\mathbf{r}^1-\mathbf{r}, t^1-t)$ and the D function is defined in Schwinger (1949); in the future $\mathbf{r}^1$ and $\mathbf{r}$ will be considered space coordinates of different electrons, although this anticipates the transformation back to configuration space.

The transition back to the Schrödinger representation can be made only when all operators in (13) refer to the same time, in which case one can perform the transformation inverse to (6)–(8). It is possible to separate off the t-dependence, which should be kept, by changing variables from t, t^1 to $t, t^1-t\equiv\tau$. One then has, using (9) and (10),

$$H_T(t)=\frac{e^2}{2}\int_{-\infty}^{\infty} d^3r^1\, d^3r \sum_{\substack{m,n,\\ p,r,}} b_m^* b_n b_p^* b_r \times \{\bar{\psi}_m(\mathbf{r}^1, t)\gamma_\mu\psi_n(\mathbf{r}^1, t)\bar{\psi}_p(\mathbf{r}, t)\gamma_\mu\psi_r(\mathbf{r}, t)\}Q_{mn}(\mathbf{r}^1-\mathbf{r}), \quad . \quad (14a)$$

where

$$Q_{mn}(\mathbf{r}^1-\mathbf{r}) \equiv \int_{-\infty}^{\infty} d\tau \exp[i(E_m-E_n)\tau]\overline{D}(\mathbf{r}^1-\mathbf{r}, \tau). \quad . \quad (14b)$$

The transformation inverse to (6)—(8) can now be performed with $\psi(\mathbf{r}, t) \rightarrow \psi(\mathbf{r})$, $\bar{\psi}(\mathbf{r}, t) \rightarrow \bar{\psi}(\mathbf{r})$, etc. The further transition from field space to configuration space is more difficult, but has been explained in

detail in Brown and Ravenhall (1951), for an analogous case. Finally, in configuration space one will have for the matrix element of H_T between given initial states n, r and final states m, p :

$$H_T(t)=e^2\int d^3r^1\,d^3r\,\bar{\psi}_m(\mathbf{r}^1)\gamma_\mu\psi_n(\mathbf{r}^1)\bar{\psi}_p(\mathbf{r})\gamma_\mu\psi_r(\mathbf{r})Q_{mn}(\mathbf{r}-\mathbf{r}^1) \quad . \quad . \quad (15)$$

(a factor 2 appears in the transition from field to configuration space). In the case $E_m=E_n$, Q can be simplified immediately, since $\bar{D}(\mathbf{r}-\mathbf{r}^1,\,\tau)$ $=(1/8\pi|\mathbf{r}-\mathbf{r}^1|)[\delta(|\mathbf{r}-\mathbf{r}^1|-\tau)+\delta(|\mathbf{r}-\mathbf{r}^1|+\tau)$ and the integration over τ leaves $Q=1/4\pi|\mathbf{r}-\mathbf{r}^1|$. The more general expression for Q will be carried further, however, since it will enter into exchange effects.

The meaning of (15) is more transparent in momentum space.

$$H_T(t)=\frac{e^2}{8\pi^3}\int d^3p_1\,d^3p_2\,d^3p_3\,\bar{\psi}_m(\mathbf{p}_2)\gamma_\mu\psi_n(\mathbf{p}_1)$$
$$\times\bar{\psi}_p(\mathbf{p}_3+\mathbf{p}_1-\mathbf{p}_2)\gamma_\mu\psi_r(\mathbf{p}_3)\frac{1}{(\mathbf{p}_2-\mathbf{p}_1)^2-(E_m-E_n)^2}. \quad . \quad . \quad (16)$$

In order to transform the interaction back into coordinate space, one must have a prescription for going around the singularity in the denominator of (16). It is well known that if only outgoing radiation is to result from real transitions, then the denominator in (16) must possess a small negative imaginary part.

One finds that Q can be expressed as

$$Q_{mn}=\frac{e^2}{4\pi}\frac{\exp\left[i\,|E_m-E_n|\,(|\mathbf{r}^1-\mathbf{r}|)\right]}{|\mathbf{r}-\mathbf{r}^1|} \quad . \quad . \quad . \quad . \quad . \quad (17)$$

Consequently, the correct interaction Hamiltonian is just

$$H_{\text{int}}=\frac{e^2}{4\pi}\frac{\gamma_\mu\gamma_\mu{}^1}{|\mathbf{r}-\mathbf{r}^1|}\exp\left[i\,|E_m-E_n|\,(|\mathbf{r}^1-\mathbf{r}|)\right], \quad . \quad . \quad (18)$$

which is just (2) in terms of Dirac's α's, when used with ψ^* instead of $\bar{\psi}$ functions.

It will be noted that for the direct diagonal matrix elements ($m=n$) and for exchange elements in case $E_m=E_n$, the term (18) reduces to

$$H_{\text{int}}=\frac{e^2}{4\pi}\frac{\gamma_\mu\gamma_\mu{}^1}{|\mathbf{r}^1-\mathbf{r}|} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (18.1)$$

which is analogous to (1) except that the retardation of the Coulomb potential is entirely absent here. The simplicity of the treatment described above is owing to the fact that one does not have to sum over bound electron states anywhere.

The writer wishes to thank Professor R. E. Peierls and Dr. R. H. Dalitz for valuable advice and criticism.

Professor G. Breit has kindly pointed out to the author in private communication that the real part of (2) is similar to an interaction derived by him in Breit (1929) p. 572, and discussed later by him. It has been

pointed out by Brown and Ravenhall (1951) that strictly speaking the basic eqn. 6 of Breit (1929) has no solutions and must be limited in some sense to processes not involving pair production. In general, these processes bring in corrections and it is not obvious that in heavy atoms these are not large. The present treatment shows that the problem can be formulated in such a way that corrections do not enter in this order of α, except in vacuum polarization terms which are not treated here. These points will be discussed more fully in a forthcoming note.

If the exponent of (2) is expanded and terms to second order are retained, then (2) goes over into the well-known Breit interaction (1), providing product wave functions of the type discussed in this note are used.

References

Breit, G., 1929, *Phys. Rev.*, **34,** 553.
Brown, G. E., and Ravenhall, D. G., 1951, *Proc. Roy. Soc.* A, **208,** 552.
Oppenheimer, J., 1930, *Phys. Rev.*, **35,** 461.
Schwinger, J., 1948, *Phys. Rev.*, **74,** 1439 ; 1949, *Ibid.*, **75,** 651.

XLIII. CORRESPONDENCE

The Double Star Effect in Nuclear Emulsion Plates Exposed to the Cosmic Rays

By G. Davis
Imperial College, London*

[Received January 28, 1952]

Li (1950) reported an excess of close star pairs over that expected on a random distribution, in plates exposed at the Jungfraujoch. The excess was of magnitude 2×10^{-2} close pairs per star, the effect extending up to a star-pair separation of $1500\,\mu$ (for $100\,\mu$ plates). This ' Li effect ' was more pronounced for plates exposed under absorbing materials. Li included 2-track stars in his analysis.

The excess is difficult to explain by any known regenerative or shower process, since rather high cross-sections for star production are required. Moreover, examination of 351 stars found in 16 cm^2 of recently searched $400\,\mu$ G.5 plates (exposed under lead) does not show the Li effect. It was therefore decided to re-analyse the two sets of plates for which Li found the maximum close-pair excess, referred to as ' ice ' and ' Pb ' plates in his paper.

The distribution of close pairs was found to be identical with Li's results. However, on subtracting 2-track stars from the analysis, the excess completely disappeared for the ' Pb ' plates (374 stars), while remaining for the ' ice ' plates (1126 stars). Both these sets of plates were searched in the early days of this laboratory, and the inexperienced searchers were very likely to miss 2-track stars. But 2-track stars may have a greater probability of being found in the neighbourhood of other stars, due to the increased alertness of the searcher in the short period following the discovery of a star (for plates with a star density of the order of 5 per cm^2). Furthermore, the difficulty of identifying 2–track stars makes their inclusion unsuitable in an analysis of high accuracy. Li was aware of this effect but thought that the searching was so efficient that it did not matter. It is unlikely that the Li effect, if it existed, should be found only if 2-track stars are included in the analysis. The 351 stars analysed by the author showed no Li effect, either when 2-track stars were included, or when they were excluded from the analysis.

It seemed very probable that the remaining excess of close star-pairs in the ice plates (2-track stars excluded) could also be due to preferential inefficient searching. One tenth of the ice plates were carefully researched

*Communicated by Sir G. P. Thomson, F.R.S.

and the original efficiency was found to be slightly under 80%, contrasting markedly with the estimate of 98% reported by Li. The excess observed could then be explained by preferential inefficient searching. The areas researched did, in fact, show a decrease in the total number of excess star-pairs. If this trend were uniform among all the ice plates, it would be sufficient to explain the remaining excess of close star-pairs.

It is concluded that the experimental data used by Li do not show that there is a double star effect.

REFERENCE

LI, 1950, *Phil. Mag.*, **41**, 1152.

The Internal Pressure of Solids

By R. O. DAVIES
Queen Mary College, University of London*

[Received January 15, 1952]

ON the basis of a Debye model Kun Huang (1951) has compared Hildebrand's internal pressure $(\partial U/\partial V)_T$, with the corresponding kinetic property defined through $p_i=\gamma E_{\text{vib}}/V-p$. He has remarked that the result depends on whether or not the zero point energy is included in E_{vib}.

The purpose of this note is to indicate how his method can be made slightly more general by considering all substances which obey the Nernst Heat Theorem and for which

$$C_V=F(T/\Theta), \qquad (\Theta=\Theta(V)). \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

Differentiation of (1) with respect to V and T yields

$$\frac{\partial C_V}{\partial V}=\frac{\gamma T}{V}\frac{\partial C_V}{\partial T}, \qquad \left(\gamma=-\frac{d\ln\Theta}{d\ln V}\right). \quad . \quad . \quad . \quad . \quad (2)$$

If $U_0(V)$ is the internal energy at $T=0$, then $U=U_0+\int_0^T C_V\,dT$ so that $(\partial U/\partial V)_T=dU_0/dV+\int_0^T(\partial C_V/\partial V)\,dT$. Carrying out the integration with the help of (2), we get

$$\left(\frac{\partial U}{\partial V}\right)_T\equiv T\frac{\alpha}{\kappa}-p=\frac{dU_0}{dV}+T\frac{\gamma C_V}{V}-\frac{\gamma(U-U_0)}{V}. \quad . \quad . \quad (3)$$

This can be simplified by first differentiating $(\partial S/\partial T)_V=C_V/T$ with respect to V so that

$$\frac{\partial}{\partial T}\left(\frac{\alpha}{\kappa}\right)=\frac{1}{T}\frac{\partial C_V}{\partial V}=\frac{\gamma}{V}\frac{\partial C_V}{\partial T}$$

---

* Communicated by the Author.

and then integrating the result. Since $\alpha=0$ at $T=0$ by the Nernst Heat Theorem, this gives

$$\frac{\alpha}{\kappa}=\frac{\gamma C_V}{V}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

Equation (4) is the Grüneisen equation. If we substitute it into (3) we get

$$p+\frac{dU_0}{dV}=\frac{\gamma(U-U_0)}{V}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

The results (3), (4) and (5) are given by Huang and provide the basis for his discussion.

REFERENCE

KUN HUANG, 1951, *Phil. Mag.*. **42**, 202.

Electrostatic Contribution to the Elastic Constants of Solids with a Diamond Structure

By E. P. WOHLFARTH

Department of Mathematics, Imperial College, London*

[Received January 30, 1952]

IN an attempt to interpret the recently measured values of the elastic constants of germanium and silicon (Bond *et al.* 1950, McSkimin *et al.* 1951), a calculation has been carried out of the electrostatic contribution to these constants using Ewald's method (1921). Such a calculation had previously been carried out by Fuchs (1935, 1936) for the face centred and body centred cubic structures. Fuchs shows that the correction to the potential energy calculated by the Wigner–Seitz method (which assumes spherical atomic cells) may be written E_l-E_s where E_l is the energy of a lattice of positive point charges in a uniform distribution of negative charge and E_s the energy of an equal number of positive point charges each surrounded by a uniform sphere of negative charge. With r_s the radius of the atomic sphere the contribution to E_s from each sphere is $-0{\cdot}9e^2/r_s$. The calculation of E_l proceeds by Ewald's method. The formula given by Fuchs (1935, p. 598) was amended so that the calculations may be carried out with the unit cell rather than the single atom as the basic unit. The summations in this formula are very rapidly convergent and the contribution per atom to E_l was found to be $-be^2/r_s$, where $b=0{\cdot}83542$. Hence $E_l-E_s=0{\cdot}06458\ e^2/r_s$. For the face centred and body centred cubic structures the values of b are 0·89586 and 0·89593 respectively, and E_l-E_s is therefore much smaller.

---

* Communicated by the Author.

Under a strain at constant volume E_s gives no contribution to the elastic constants. The contribution from E_l was calculated by inserting the values of the components of the lattice vectors and the reciprocal lattice vectors as functions of a suitably chosen strain into the summation formula for E_l and differentiating twice with respect to the strain as this tends to zero. The contributions to the elastic constants are

$$\tfrac{1}{2}(c_{11}-c_{12}) = -0{\cdot}1544\, Z_{\text{eff}}^2 e^2/2a \text{ per atom},$$

$$c_{44} = \quad 0{\cdot}2526\, Z_{\text{eff}}^2 e^2/2a \text{ per atom},$$

where a is the edge length of the unit cell. These values may be compared with those given by Fuchs (1936) for the face centred and body centred cubic structures. The surprising result is that $\frac{1}{2}(c_{11}-c_{12})$ is negative. This result was checked by actually evaluating E_l as a function of strain in this case, the energy, strain curve having indeed a maximum at the 'equilibrium' position. It may be noted that the reason for the negative sign is closely related to the peculiar lattice structure of the diamond lattice and to the low coordination number.

In order to investigate this quasi-geometrical effect further, a calculation of the elastic constants was carried out assuming central forces to act between every atom in the diamond lattice. It was assumed that the potential energy per atom pair separated by a distance r, $w(r)$, is of the Grüneisen form (1926),

$$w(r) = -ar^{-m}+br^{-n}, \quad n>m.$$

For the diamond structure *with nearest neighbours only interacting* it is easy to show that

$$\tfrac{1}{2}(c_{11}-c_{12}) = \tfrac{2}{3}\, r_0 w'(r)_{r=r_0} \text{ per atom},$$

$$c_{44} = \tfrac{4}{9}\, r_0 w'(r)_{r_0} + \tfrac{2}{9}\, r_0^2 w''(r)_{r_0} \text{ per atom},$$

where r_0 is the equilibrium nearest neighbour distance. Hence $w'(r)_{r_0}=0$, $\frac{1}{2}(c_{11}-c_{12})=0$, while $c_{44}>0$. The calculation has been extended to take account of next nearest neighbour interaction. The equilibrium value of r now no longer coincides with the minimum of $w(r)$; the final expressions for the elastic constants are too complicated to reproduce here, but a plot of $\frac{1}{2}(c_{11}-c_{12})$ against the index m for a range of values of n shows that for all reasonable values of the indices (Grüneisen *loc. cit.*) $\frac{1}{2}(c_{11}-c_{12})$ is negative.

In order to investigate whether the diamond lattice is stabilized by a Fermi energy contribution a calculation similar to that of Leigh (1951) was carried out for the electrons just filling the {220} zone, which contains 4 electrons per atom. It was assumed that the energy, wave vector relation is of the form $E=\alpha_0 h^2k^2/8\pi^2 m$, as in Leigh's calculation for the full zone of aluminium. This assumption is difficult to justify for germanium and silicon, and it is indeed found that the Fermi energy contribution fails to stabilize the diamond lattice under the '$\frac{1}{2}(c_{11}-c_{12})$'

strain, although, with reasonable values of the parameters α_0 and Z_{eff} (Leigh *loc. cit.*), agreement may be obtained with the observed values of c_{44}.

Finally, for germanium and silicon the non-Coulomb interactions between nearest neighbours are very weak owing to the large clearance between the ion cores (Jones 1949). Hence this contribution to the elastic constants is here negligible.

It must be concluded that the solids are stabilized by some contribution arising from a non-uniform distribution of electronic charge, such as would reside in the directed valency bond approach of Slater and Pauling. Work on this and related topics is proceeding.

References

Bond, W. L., *et al.*, 1950, *Phys. Rev.*, **78**, 176.
Ewald, P. P., 1921, *Ann. Phys. Lpz.*, **64**, 253.
Fuchs, K., 1935, *Proc. Roy. Soc.* A, **151**, 585 ; 1936, *Ibid.*, **153**, 622.
Grüneisen, G., 1926, *Handbuch der Physik*, **10**, 1.
Jones, H., 1949, *Physica*, **15**, 13.
Leigh, R. S., 1951, *Phil. Mag.*, **42**, 139.
McSkimin, H. J., *et al.*, 1951, *Phys. Rev.*, **83**, 1080.

The α-activities of the Two 'Long-lived' States of $^{211}_{84}$Po

By N. Feather, F.R.S.
The University, Edinburgh*

[Received February 4, 1952]

In a recent note (Feather 1951) I have discussed the problem posed by the properties of the isomeric state of $^{210}_{83}$RaE which decays almost exclusively by α-disintegration. More recently Spiess (1951) has reported measurements which show that the α-active AcC′ of classical radioactivity, which is identical with the daughter product of the capture mode of disintegration of $^{211}_{85}$At, is the species $^{211}_{84}$Po in an isomeric state of about 0·3 MeV energy of excitation. Experiment indicates that in the decay of $^{211}_{85}$At the ground state $^{211}_{84}$Po is produced less frequently than once in 10^4 capture disintegrations (Spiess *loc. cit.*), and that this state is produced less frequently than once in 10^2 β-disintegrations of $^{211}_{83}$AcC (Rutherford, Wynn-Williams and Lewis 1931). According to Spiess, also, an upper limit of 10^{-4} can be set to the ratio of the probabilities of direct de-excitation and of α-disintegration of $^{211}_{84}$Po* (0·52 sec.). All this information requires that there is a large difference of spin (possibly 5 units, or even more) between the ground state $^{211}_{84}$Po and the isomeric state $^{211}_{84}$Po*—a conclusion which poses a problem in respect of the α-activities of these two states which is closely similar to that previously discussed in respect of the two states of $^{210}_{83}$Bi

* Communicated by the Author.

(RaE and its isomer). The outcome of the previous discussion was the inference that the two states of $^{210}_{83}$Bi differed markedly in the 'difficulty of formation' of the outgoing α-particle, the 'difficulty factor' being considerably less for the isomeric state than for the ground state of that nucleus. It will be seen, from the considerations which follow, that a similar difference, though in the opposite sense, must be assumed to characterize the two 'long-lived' states of $^{211}_{84}$Po.

Here it is necessary to be a little more explicit concerning the method used for deducing values of the spin change involved in α-disintegration. This question is treated systematically in a monograph in course of publication (Feather 1952). There reasons are given for supposing that the degree of forbiddenness due to a change of spin of j quantum units can be represented by a factor F in the disintegration constant λ such that

$$\log_{10} F = -C(Z-2)^{-1/2}\, \rho_0^{-1/2} j(j+1). \quad . \quad . \quad . \quad . \quad . \quad (1)$$

Here ρ_0 is the nuclear radius (in units of 10^{-12} cm), Z is the atomic number of the disintegrating nucleus, and $C = 0{\cdot}70 \pm 0{\cdot}03$ (in appropriate units). This is effectively assuming the form of the forbiddenness factor originally given by Gamow, but replacing his multiplying constant 0·372 (in our units) by the empirically determined value of C quoted above. It is also consciously disregarding the more ambitious treatment of Preston (1947), which Spiess has preferred in his discussion of the problem, on the grounds that the one-body model is surely an inadequate basis on which to discuss in detail 'the formation and escape of an α-particle from a heavy nucleus' (Preston *loc. cit.*, p. 874). From our point of view 'difficulty of formation' is considered separately, after what are regarded as the more regular effects of spin-change have been taken count of. Clearly, the success of the treatment depends on the presumption—which a general survey of results appears to confirm—that in most cases the difficulty factor varies little as between the different modes of α-disintegration of a single body.

On the basis of eqn. (1), and on the assumption that the same difficulty factor which characterizes the α-disintegrations of the isotopes of polonium of even A (mass number) also applies to the α-disintegration of $^{211}_{84}$Po (25 sec.), we obtain $j = 5$ for this disintegration. This is in agreement with the conclusions of Spiess and suggests 11/2 quantum units as the most probable value for the spin of $^{211}_{84}$Po. On a comparable basis a spin of 5/2 has been suggested for $^{211}_{83}$AcC (Feather 1952). These two assignments lead us to predict that the production of $^{211}_{84}$Po cannot be much more than 10^{-3} times as probable as the observed production of $^{211}_{84}$Po* in the β-disintegration of $^{211}_{83}$AcC (see Feather and Richardson 1948). This again is in full agreement with observation (*v. sup.*); moreover the agreement could not be maintained on the present basis of interpretation if it were supposed that the difficulty factor for the α-disintegration of $^{211}_{84}$Po is markedly greater than the difficulty factor which characterizes the α-disintegrations of the polonium isotopes of even A.

The situation is entirely different if the above assumptions are made in respect of the α-disintegration of $^{211}_{84}$Po*. We then obtain $j=4$, 4, 3 and 3 (or 2) as most probable values for the α-disintegrations of total disintegration energy 7·58, 7·03, 6·70 and 6·46 MEV, respectively (Neumann and Perlman 1951). If this result is accepted, and if the same final state is involved in the 7·58 MEV mode of $^{211}_{84}$Po* as in the 7·28 MEV disintegration of $^{211}_{84}$Po, then the spins of the 'long-lived' initial states cannot differ by more than 1 unit. This conclusion is entirely unacceptable, chiefly, as already mentioned, in view of the unobservably small rate of direct de-excitation of the upper state. As Spiess has concluded, we require j values for the α-disintegration modes of $^{211}_{84}$Po* very much smaller than those quoted above. On the basis of interpretation here adopted this implies the assumption that the difficulty factor is much greater for the α-disintegration $^{211}_{84}\text{Po}^* \rightarrow ^{207}_{82}\text{Pb}$ than for the disintegration $^{211}_{84}\text{Po} \rightarrow ^{207}_{82}\text{Pb}$.

It may well be that our basic assumptions represent only a crude approximation in that in fact it may be generally impossible to separate out a j-independent difficulty factor in α-disintegration, but naively accepting these assumptions we have been led to conclude (for $^{210}_{83}$Bi and $^{211}_{84}$Po) that when isomeric states are compared (that is when widely different j values are in question) considerably different values of the difficulty factor are required. In this connection two further remarks only need be made here. The first is that with both species the difficulty factor is greater for the state of lower spin ($^{210}_{83}$Bi in the one case and $^{211}_{84}$Po* in the other), and the second remark, quite unconnected with the first, is that the upper state $^{210}_{83}$Bi* appears to be characterized by a difficulty factor more normal for the ground state of an isotope of polonium ($Z=84$) than for the ground state of an isotope of bismuth ($Z=83$), whereas the upper state $^{211}_{84}$Po* seems to behave in respect of α-formation more like a ground state bismuth isotope than a ground state polonium isotope. The two species which we are comparing have, of course, the same neutron number ($N=127$) and proton numbers differing by one.

References

FEATHER, N., 1951, *Phil. Mag.*, **42**, 568; 1952. *Nuclear Stability Rules*, C.U.P. (in course of publication).

FEATHER, N., and RICHARDSON, H. O. W., 1948, *Proc. Phys. Soc.*, **61**, 452.

NEUMANN, H. M., and PERLMAN, I., 1951, *Phys. Rev.*, **81**, 958.

PRESTON, M. A., 1947, *Phys. Rev.*, **71**, 865.

RUTHERFORD (Lord), WYNN-WILLIAMS, C. E., and LEWIS, W. B., 1931, *Proc. Roy. Soc.* A, **133**, 351.

SPIESS, F. N., 1951, *American Report, UCRL*, 1494.

Erratum

FEATHER, N., 1951, *Phil. Mag.*, **42**, 568.

p. 569 *for* '(ii) the β-disintegration . . .' *read* '(ii) the α-disintegration . . .'.

Experiments on the Electrical Resistivity of Metals

By D. K. C. MacDonald
National Research Council, Ottawa*

[Received February 10, 1952]

It has been found possible to adapt a Collins Helium Liquefier so that electrical conductivity measurements may be carried out continuously from room temperature down to the liquid helium temperature range in a single experiment with very satisfactory accuracy. The metallic specimen is mounted alongside a helium gas thermometer of the general type used earlier by Simon for low temperature work. The specimen and thermometer bulb are fixed inside a simple glass-sheathed copper vessel designed to provide a uniform temperature enclosure within the low-temperature chamber of the liquefier. A careful choice of the thermometer bulb size in relation to the dead-space volume enabled an adequate relative accuracy of thermometry to be maintained throughout.

Results of very considerable interest have been found already in the case of the alkali metals, and we wish here to report briefly some measurements on rubidium. Fig. 1 shows the variation of electrical resistance observed together with a number of points, detailed in the caption, taken in fixed temperature baths which check the accuracy and consistency of the temperature measurements. It is immediately clear that the resistance becomes anomalous above about 170° K and that, in the region above about 200° K, the metal is in a different state from that below the former temperature. It is immediately satisfying to observe that these measurements now explain the very remarkable variation of the apparent Debye characteristic temperature, Θ, deduced by MacDonald and Mendelssohn (1950) from measurements below 90° K using the observed room-temperature value as a reference value. When, however, the calculation is based on an extrapolation of the lower part of our curve, or by using the slope of the resistance curve at each point, an essentially *normal* variation of Θ with temperature is found in this region as shown in table 1.

Table 1

T(°K)	10	20	50	90
Θ_1(°K)	58	75	120	192
Θ_2(°K)	54	64	71	81

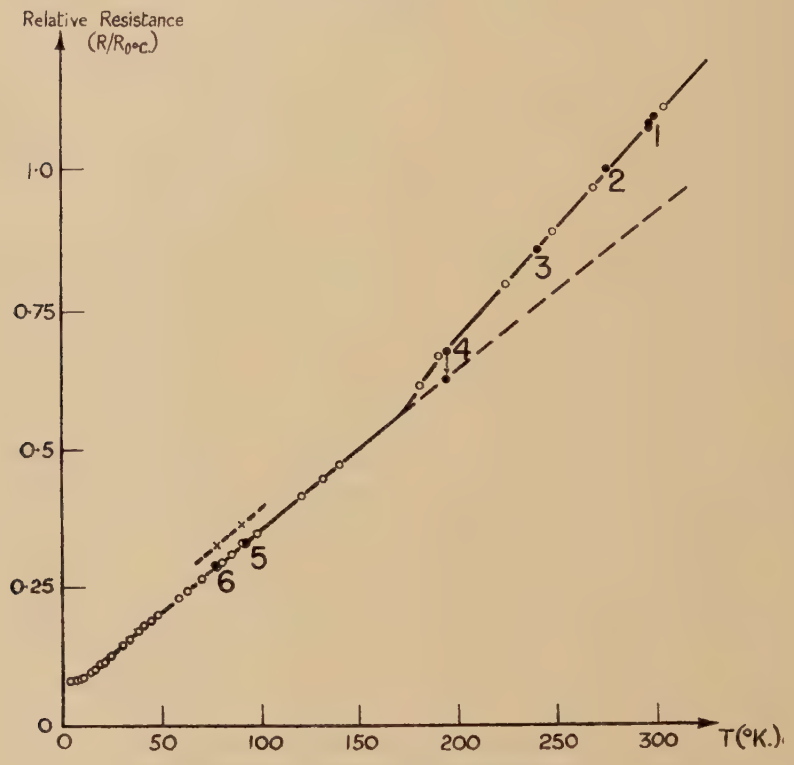
Θ_1 : As deduced from full curve of fig. 1, using *observed* room-temperature value as reference point.

Θ_2 : As deduced using reference point from extrapolation of lower part of curve.

* Communicated by the Author.

The precise nature of the transition around 180°–190° K is not yet clear, but it is interesting to find that measurements in a CO_2 bath show a time-dependent variation as indicated in fig. 1, and also that by very rapid cooling of the specimen in pre-pumped liquid nitrogen, something akin to a ' freezing-in ' of the transition may be achieved as is also shown in fig. 1 by following observations at liquid nitrogen and oxygen

Fig. 1



Electrical resistance of Rubidium.

O : Observations on cryostat experiment.

● : Observations in fixed-temperature baths
- (1) Room-temperature.
- (2) Ice point.
- (3) Liquid ammonia.
- (4) Solid carbon dioxide (arrow shows range of time-dependent resistance).
- (5) Liquid oxygen.
- (6) Liquid nitrogen.

× : Observed points in liquid oxygen and nitrogen after very rapid cooling.

temperatures. It should be remarked that re-heating above the transition region and subsequent *slow* cooling reproduces once more the full curve. It may be mentioned in passing that we have now therefore an explanation of reports at various times that an alkali metal has been

found to 'anneal' (i.e. the resistance is observed to diminish over a sufficiently long period of time) at temperatures around the liquid oxygen region.

In view of the experiments of Simon and Vohsen (1928) who determined the crystal structure of alkali metals condensed at ~90° K and found no variation whatsoever from a body-centred cubic structure, it seems extremely improbable that a crystallographic transformation (e.g. to face-centred cubic structure) is responsible for this behaviour; it is intended, however, with the co-operation of the X-ray analysis group to re-examine this possibility. In the absence of any crystallographic modification, it seems certain that some change in electronic configuration occurs, perhaps similar to that suggested by Fermi and Sternheimer (Sternheimer 1950) for the volume transition in Cs under pressure found by Bridgman. We have found analogous behaviour in K and Cs, while Li and Na exhibit *no* such anomaly. The K, Rb and Cs atoms in their normal state have vacant *d* and/or *f* levels before the valence electron, while all levels are occupied in the Li and Na atoms.

Finally, the measurements below ~20° K have again confirmed the occurrence of sharp 'kinks' such as first reported at low temperatures in K and Cs by MacDonald and Mendelssohn (*loc. cit.*).

A full survey of the alkali metals is being completed, together with other physical measurements of interest.

References

MacDonald, D. K. C., and Mendelssohn, K., 1950, *Proc. Roy. Soc.* A, **202**, 103.
Simon, F., and Vohsen, E., 1928, *Zs. Phys. Chem.*, **133**, 165.
Sternheimer, E., 1950, *Phys. Rev.*, **78**, 235.

Growth Spirals on Magnesium Crystals

By A. J. Forty
H. H. Wills Physical Laboratory, University of Bristol*

[Received March 1, 1952]

Growth spirals, predicted by the theory of crystal growth (Frank 1949, 1952), have been found on many different substances. This letter records the observation of spirals on the surfaces of magnesium crystals; this is the first direct indication of the presence of screw dislocations in a simple metal crystal, though screw dislocations of large Burgers vector have been seen in chemically deposited gold (Amelinckx 1952). It is thought that the growth layers are either one or two atoms in thickness (i.e. steps of 2·6 A or 5·2 A) depending on whether magnesium grows in the [0001] direction by layers containing a half or a whole unit

* Communicated by the Author.

cell. The detection of such small growth layers is normally beyond the limit of existing microscopic techniques, but in the present observation the steps are made visible by an extraordinary ' decorating ' phenomenon. It has already been established that a decoration of the edges of monolayers by a concentration of small etch pits on beryl (Griffin 1951) and by ridges of material on some crystals of silicon carbide (Verma 1951, Vand 1951) accounts for their remarkably high visibility.

The magnesium crystals studied were thin hexagonal flakes, about 1 to 2 mm across, having well-developed shiny basal (0001) surfaces, obtained by sublimation. Turnings of pure magnesium (99·99%) were placed in a nickel foil liner in a nickel tube, heated by an electrical winding to a peak temperature of 620° C (the melting point is 659° C), and held at this temperature in a very slow stream of argon. In the course of about 12 hours the crystals appeared on some of the cooler turnings. They were mounted on plasticine for observation under the reflection phase contrast microscope. Initially no step structure can be seen on the surfaces even under perfect phase contrast adjustment ; if there are any growth steps on the surfaces they are therefore thought to be less than 20 A high. Interference fringes produced between the metal surface and a glass flat are observed to traverse the whole surface apparently without deflection, which indicates that there are no steep growth hills. Therefore any small growth steps on the surface are likely to be sufficiently well spaced for a lateral resolution. This latter conclusion is confirmed by the appearance of extremely faint growth steps after the crystal has been in contact with the plasticine for about one hour. The visibility of the steps can be improved by photography and growth hills have been found having the expected spiral terrace form based on screw dislocations with their Burgers vectors normal to the basal plane. After exposure to air and plasticine for about 3 hours the crystal surface becomes ' speckled ' and it is difficult to detect the growth steps. Often, after about 24 hours' exposure, the further development of the surface renders the growth steps highly visible as bright lines on a darker, speckled background.

Figure 1 (Plate XXI*) is a micrograph of an (0001) surface of a magnesium crystal after an exposure to the developing action of the plasticine for about one hour (the contrast of this photograph has been increased by two stages of copying, to make reproduction possible). It shows a growth hill consisting of two parallel spiral terraces of many turns with two similar screw dislocations at its centre. Figure 2 (Plate XXI) is a micrograph of the same surface after development for 24 hours. The steps are now more visible, especially in the middle of the surface where the development is not so complete. Figures 3 and 4 (Plate XXI) are micrographs of other crystals which have been exposed to the developing action of plasticine for about 24 hours ; the spiral step-line is clearly

---

* For plate see end of issue.

visible. It is not yet certain whether the growth layers are one or two atoms in thickness. On all the crystals examined the layers are rounded, which is expected for growth from the vapour phase, but sometimes they tend towards a trigonal shape.

Experiments have shown that this decoration of the layer edges is dependent on an optimum exposure of the crystal surface to both air and plasticine. It is always found that the development spreads inwards from the edge of the crystal. Although the phenomenon is not yet fully understood, it is thought that an oily film from the plasticine creeps over the crystal surface. It is probable that oxide on the metal surface is also involved, possibly combining with fatty acids from the plasticine, and this oxide may be thicker at the edges of the growth layers. This could produce along the steps effective ridges, to which phase-contrast microscopy is most sensitive.

The phenomenon is being investigated further and the results of this research, together with more observations of the presence and properties of screw dislocations in magnesium crystals, will be given in a future paper.

I should like to thank Dr. F. C. Frank for his continued interest in this work, and the Department of Scientific and Industrial Research for a grant.

References

Amelinckx, S., 1952, *Comptes Rendus*, **234,** 113.
Frank, F. C., 1949, *Discussions of the Faraday Society*, **5,** " Crystal Growth ", 49 ; 1952, *Advances in Physics* (*Phil. Mag. Supplement*), **1,** 91.
Griffin, L. J., 1951, *Phil. Mag.*, **42,** 775.
Vand, V., 1951, *Phil. Mag.*, **42,** 1384.
Verma, A. R., 1951, *Phil. Mag.*, **42,** 1005.

XLIV. *Notices of New Books and Periodicals received*

Classical Mechanics. By H. Goldstein. (Addison-Wesley Press, Inc., Cambridge, Mass., 1951.) [Pp. xii+399.] Price $6.50.

This is an excellent book, tailored to the needs of physicists and laying stress on those aspects of classical mechanics which provide the background of modern physics.

The first five chapters treat Lagrange's equation and the associated variational principle, central forces and rigid body motion. Chapter 6 presents the elements of the special theory of relativity, shows how the equations of mechanics must be altered to conform with it and expresses them in Lagrangian form. The next three chapters deal with Hamilton's equations and transformation theory (canonical transformations, Hamilton–Jacobi equations, Poisson brackets). Chapter 10 treats the theory of small oscillations.

The final chapter discusses the Lagrangian and Hamiltonian formulation of the equations of continuous systems (fields) and their derivation from variational principles. It is to be hoped that the author will be able to fulfil a half-promise in his preface and in a future edition treat in their classical setting such topics as the energy-momentum tensor and the related conservation theorems. They are of interest, for example, in the theory of elasticity, where the canonical energy-momentum tensor (as opposed to the ordinary stress tensor) is useful in calculating the forces on dislocations and other elastic singularities.

The author indicates the points where classical and quantum mechanics make contact and develops mathematical methods (e.g. the manipulation of matrices) with an eye to their wider applications. Each chapter ends with a bibliography enlivened with shrewd comments. J. D. E.

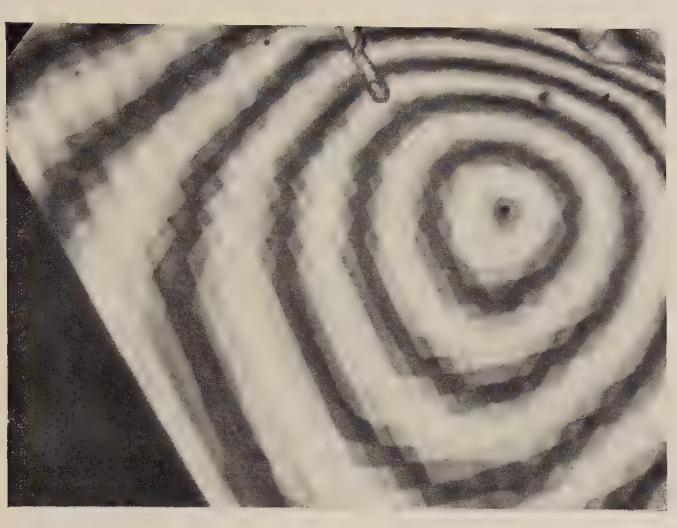
The Structure of Physical Chemistry. By Prof. C. Hinshlewood. (Oxford University Press.) Price 35*s*.

The rapid advances of the physical sciences oblige the research worker to study an ever decreasing portion of Nature with an ever increasing thoroughness and so ' to learn more and more about less and less '. Not the least of the dangers of such specialization is that the scientist tends to lose touch with the fundamental bases of his own subject and to see it from the limited viewpoint of his own specialization.

It is then most welcome and refreshing to find a book dealing with the whole field of physical chemistry, at an advanced but by no means mathematically difficult level, in which the emphasis is laid not on exhaustive treatments of any topics, but on the unity and logical continuity of the whole subject, on the interdependence and relative importance of the many theories and concepts upon which the subject freely draws and on the limitations and ultimate physical basis and content of those theories and concepts. The author makes no claim that his book in any way replaces the usual text books on the subject; but this is a little modest for many standard ' results ' are derived as fully as in many text books and often with more clarity. It is a well written and easily readable book and is to be recommended to all students of the physical sciences as a most stimulating critical survey of and commentary on the subject of physical chemistry. The addition of a bibliography of standard texts and of references, at least to some of the more imported original papers, would be helpful for students who might use the book as an introductory survey of the subject. A. D. Lec.

[*The Editors do not hold themselves responsible for the views expressed by their correspondents.*]

Fig. 5



(×650) Internal interferogram of crystal plate with step-structure on both surfaces, centred on same dislocation group.

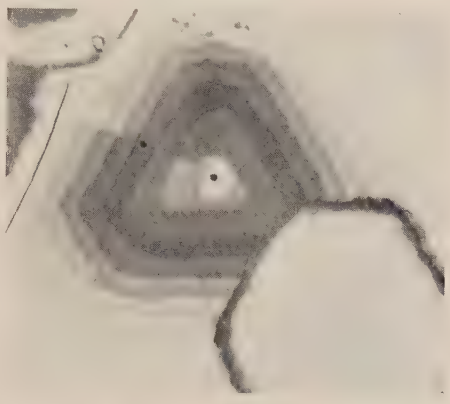
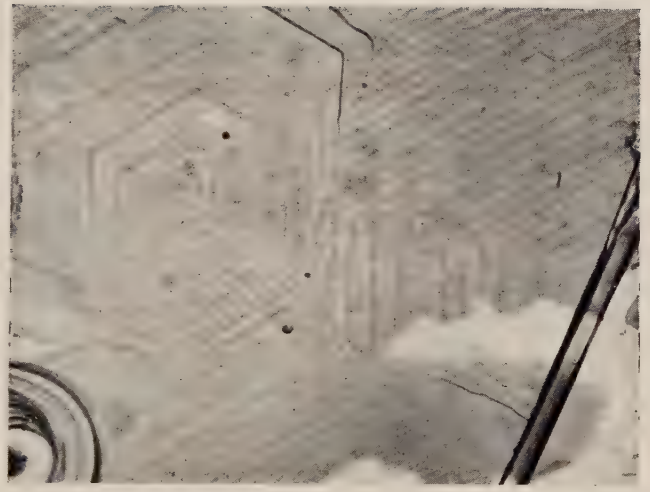


Fig. 6
(×650) Internal interferogram of crystal plate containing one screw dislocation group, with growth on one surface only, (step-height 96 ± 1 A).

Fig. 7

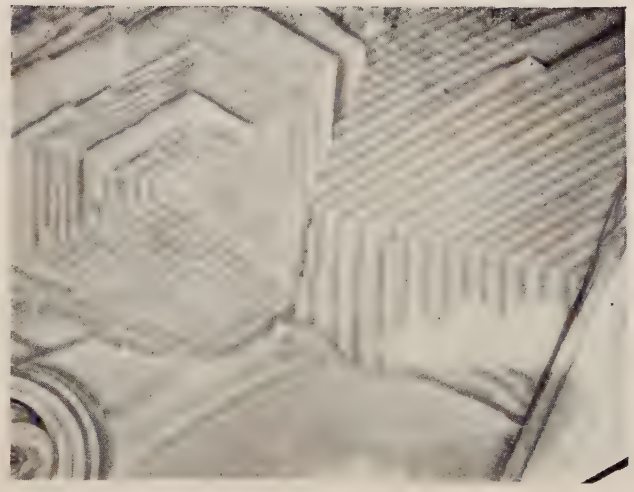
(×650) Internal interferogram of crystal plate containing two co-operating screw dislocation groups of opposite hand, with growth on one surface only, (step-height=246±0.5 A).

Fig. 8



(×650) Bright field, vertical illumination.

Fig. 9



(×650) Reflection, phase contrast illumination.

Fig. 10

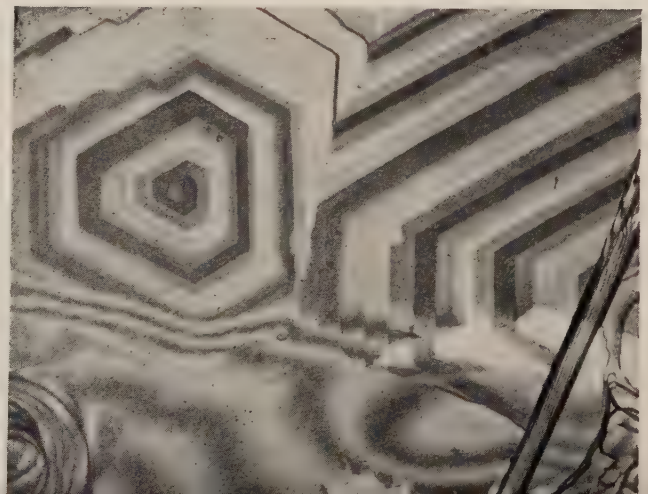


Fig. 11

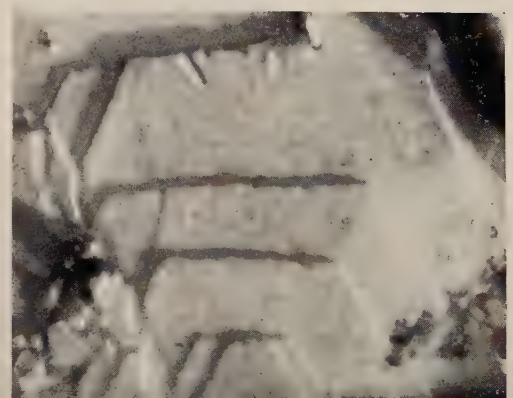
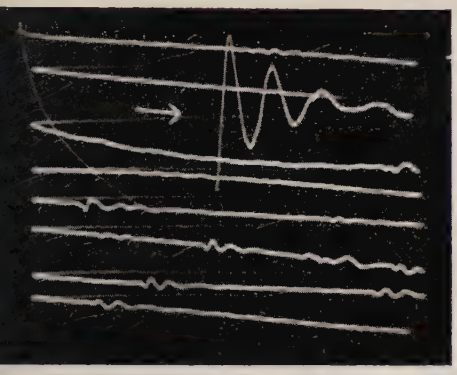


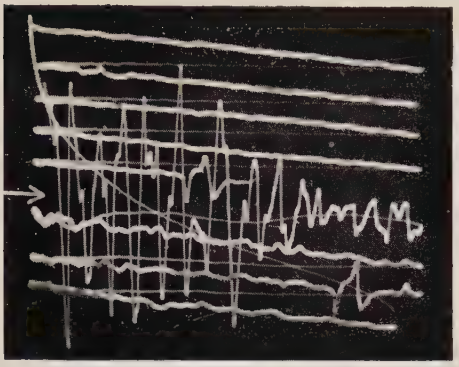
Fig. 1



0·1 v/m

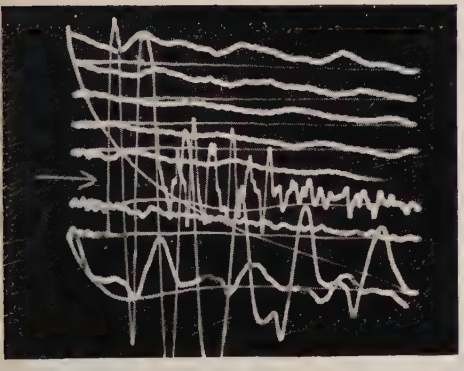
2·05 msec

Fig. 2



2·05 msec

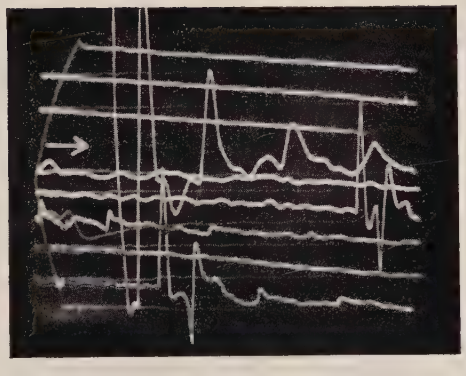
Fig. 3



0·1 v/m

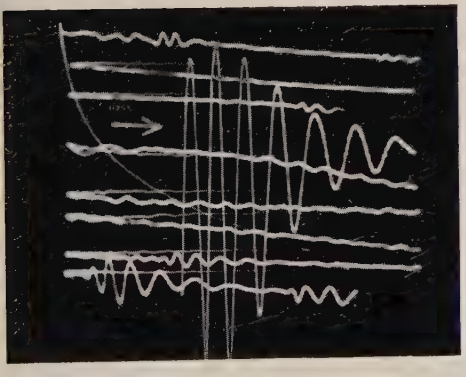
2·05 msec

Fig. 4



2·05 msec

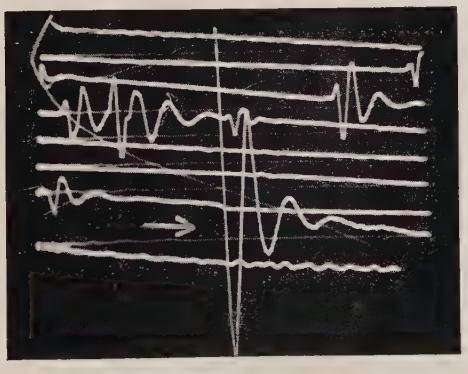
Fig. 5



0·1 v/m

2·05 msec

Fig. 6



2·05 msec

Fig. 7

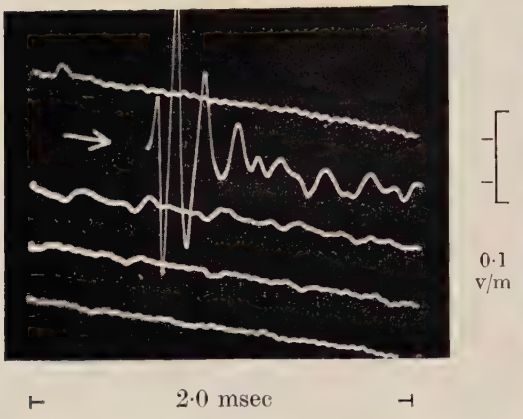


Fig. 8

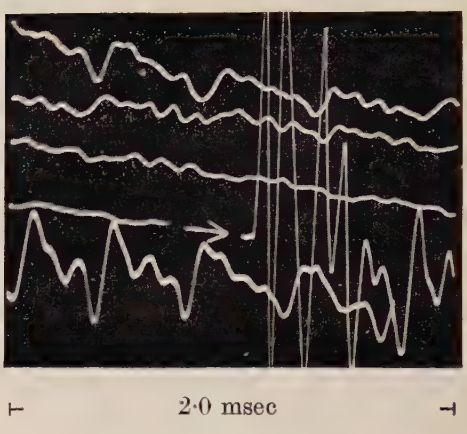


Fig. 9

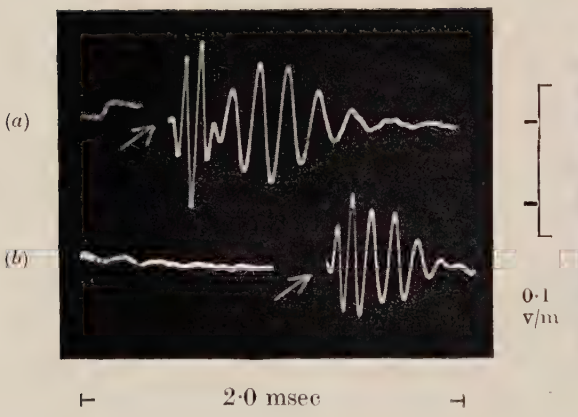


Fig. 10

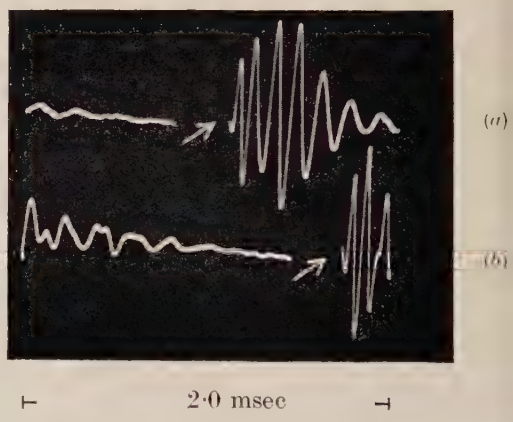


Fig. 11

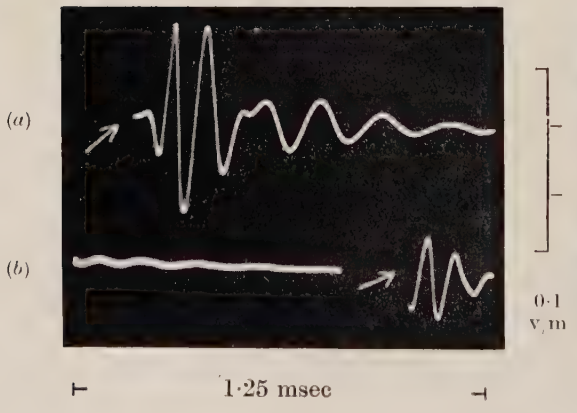


Fig. 12

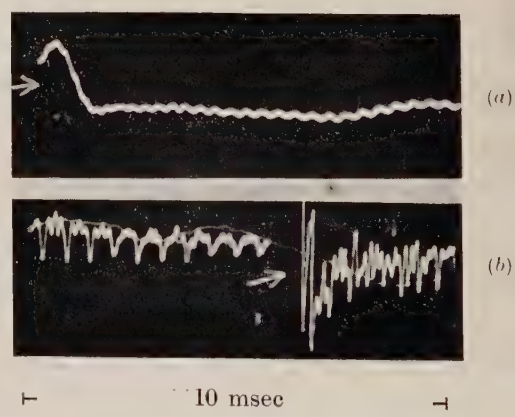
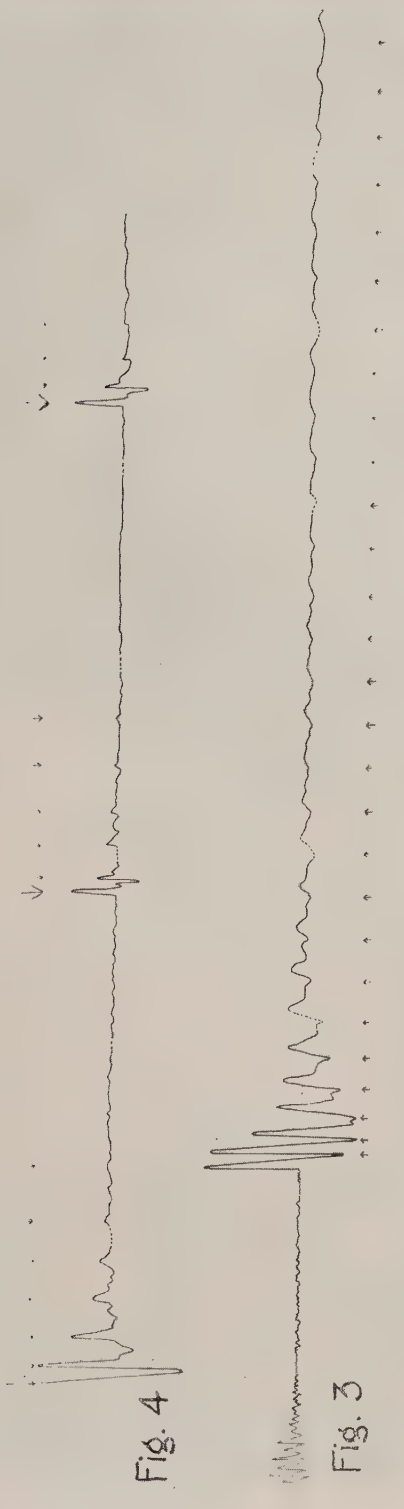
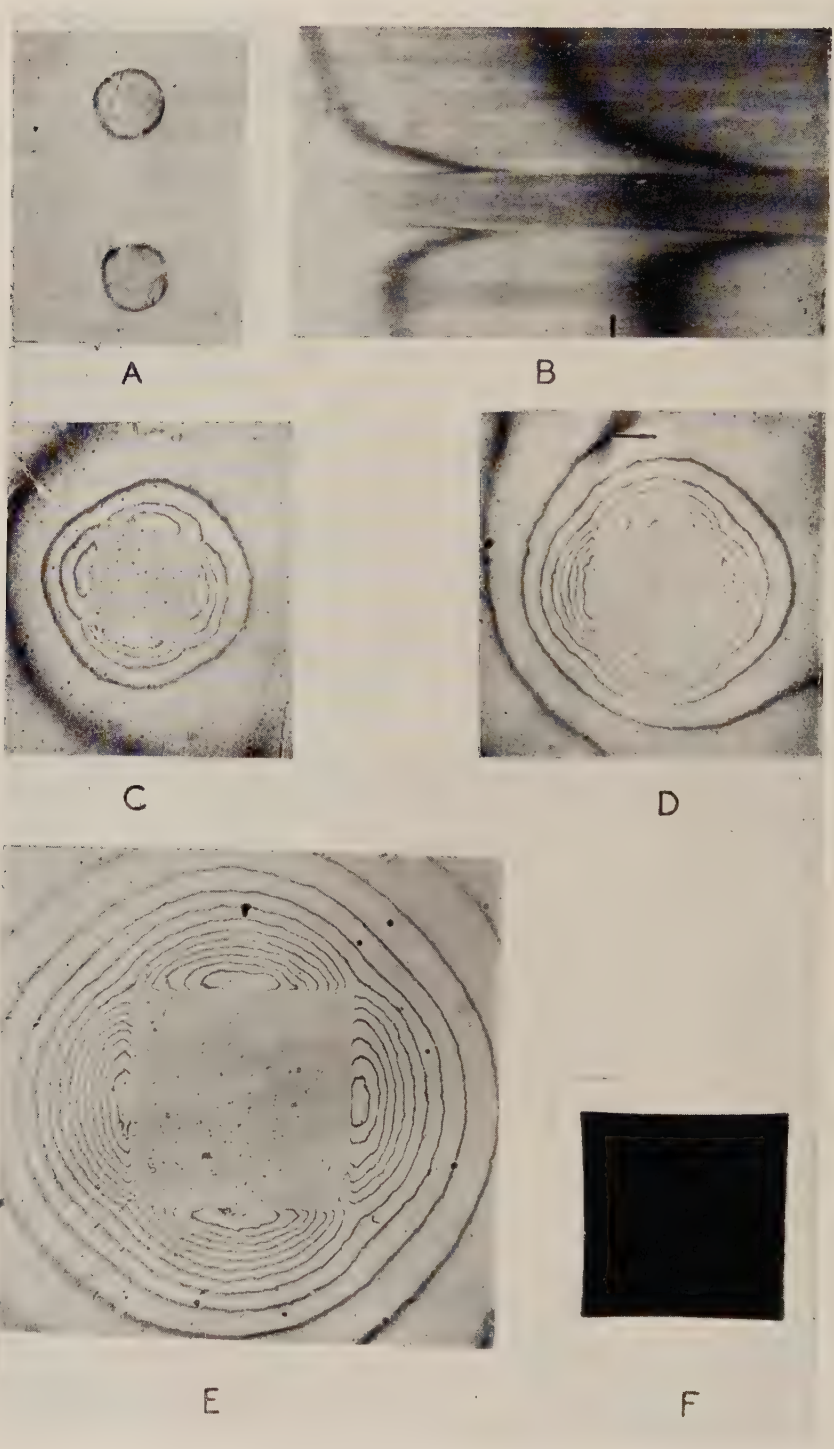
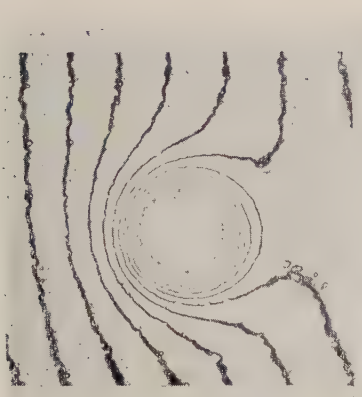


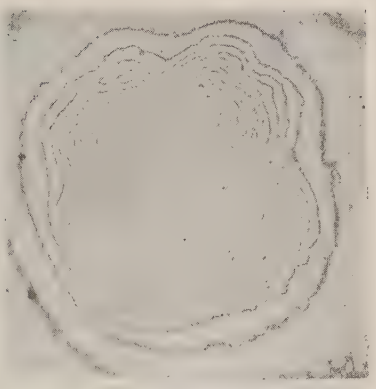
Fig. 14



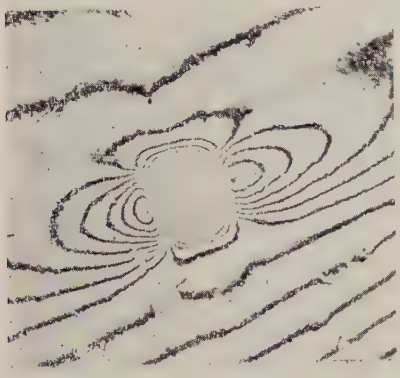

A
B
C
D
E
F



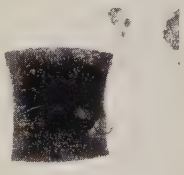
A



B



C



D

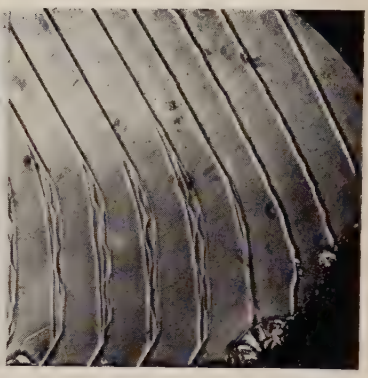
E

Fig. 1

×60

Fig. 3

Fig. 2



×100

Fig. 4

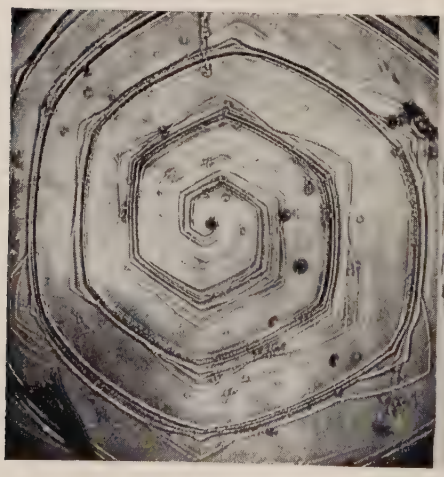
×60

Fig. 5

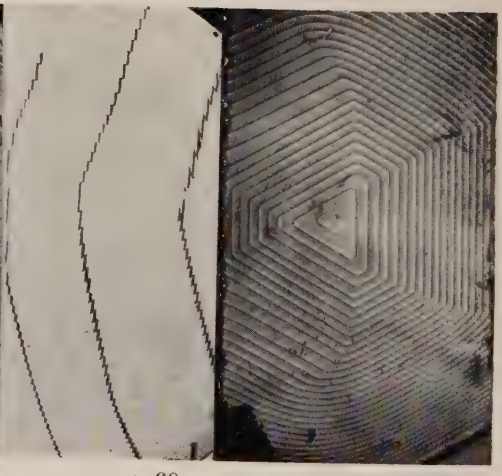
×60

Fig. 6

Fig. 7



×100



×60 ×60

Fig. 8

×90

Fig. 10

Fig. 9

×130

Fig. 11

Fig. 12

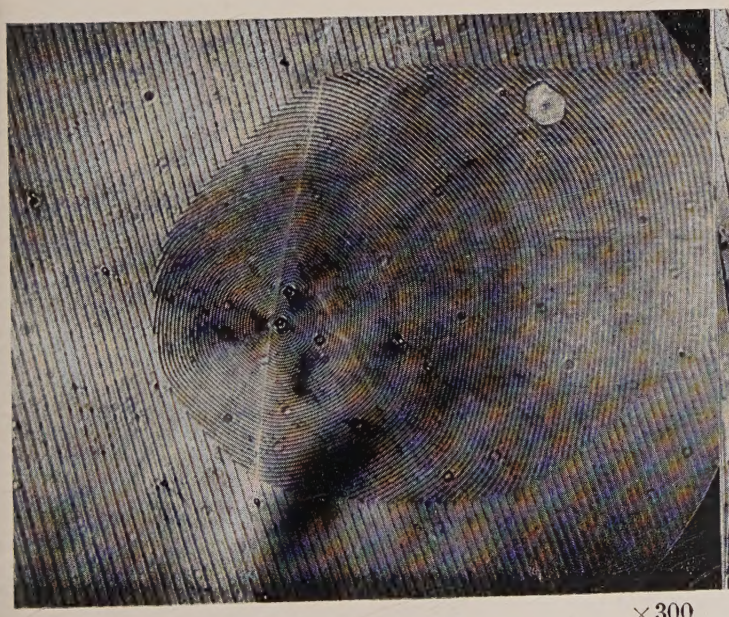
×150

Fig. 13

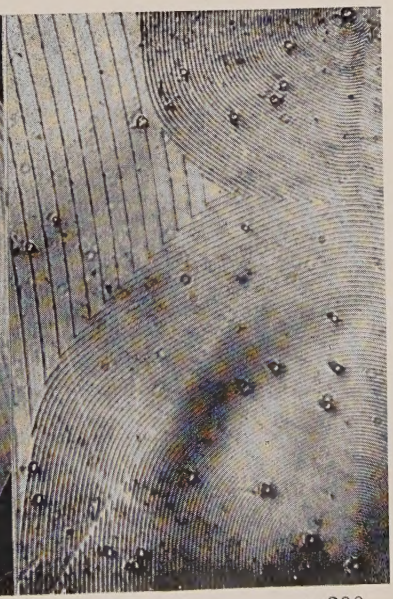
×90

×90

Fig. 14

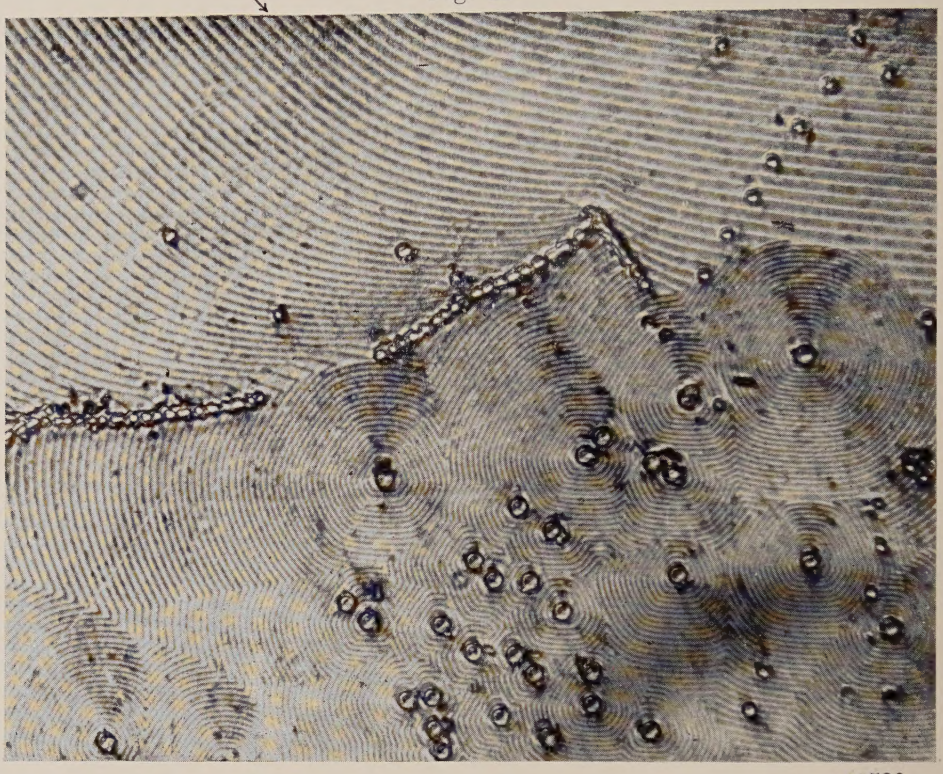


×300



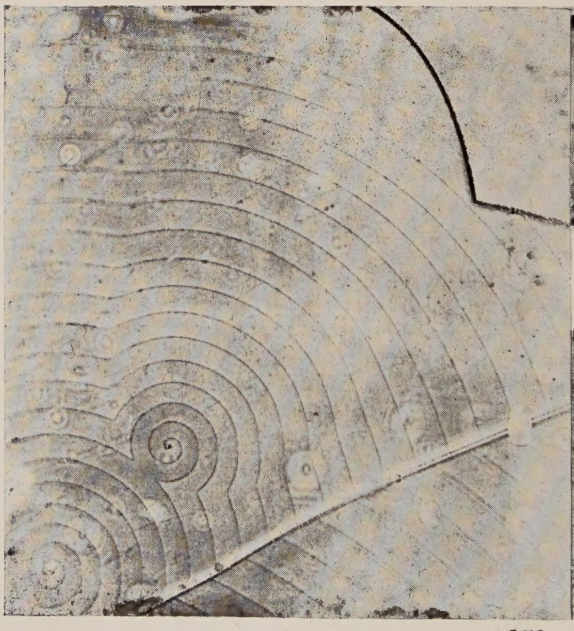
×300

Fig. 15



×580

Fig. 16



×150

Fig. 17

×90

Fig. 18

×300

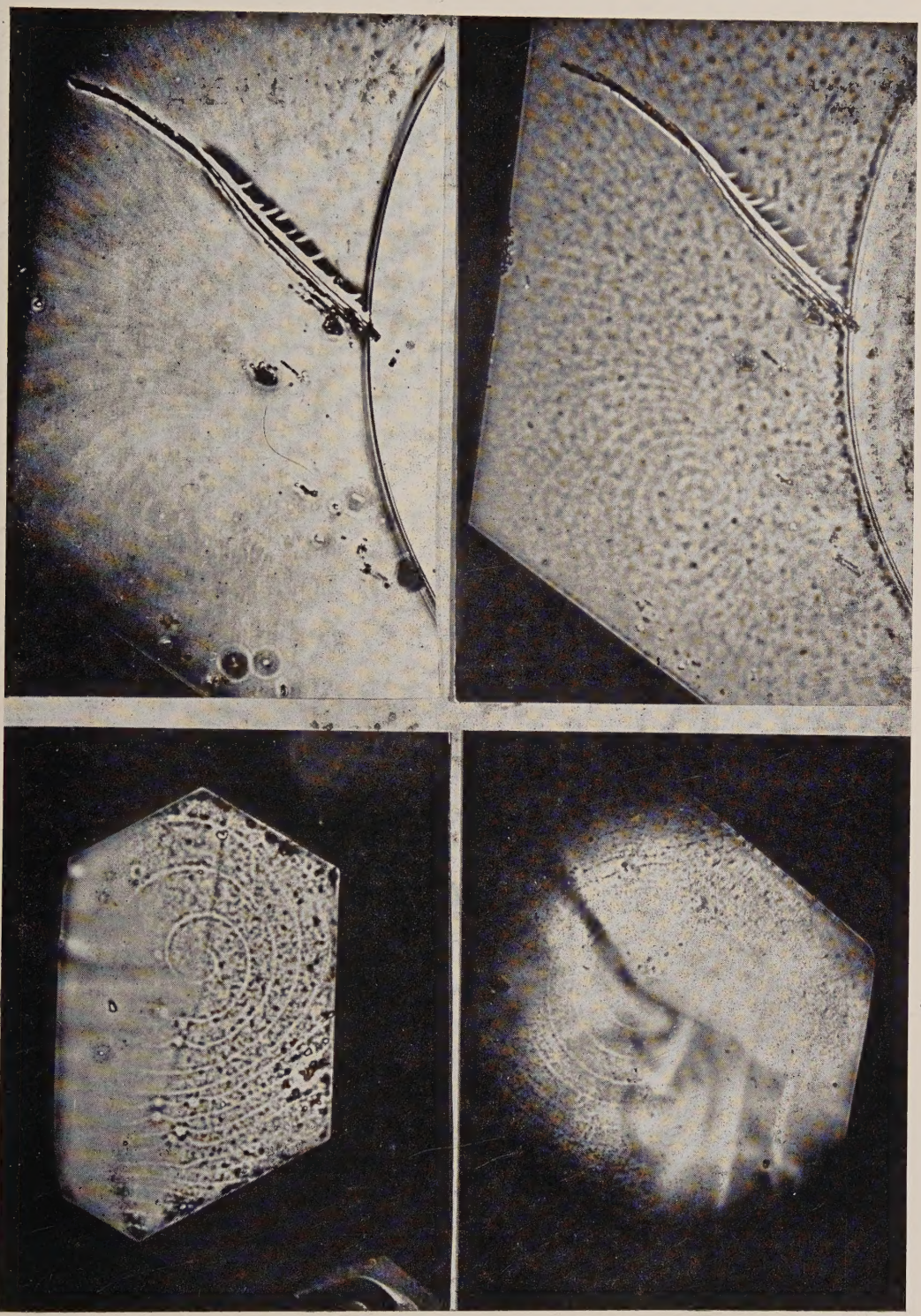


Fig. 1.—Magnesium crystal photographed under reflection phase contrast illumination after exposure to air and plasticine for one hour. ×325.

Fig. 2.—Same crystal as in fig. 1 but photographed after exposure to air and plasticine for 24 hours. Reflection, phase contrast illumination. ×325.

Figs. 3 and 4.—Micrographs of magnesium crystals showing how the ridges have been developed along the edges of the spiral growth layers after exposure to air and plasticine for 24 hours. Reflection, phase contrast illumination. ×325.